

Evidence Synthesis and targeting further research for adherence and stratification in health economic evaluations



Claire Louise Simons

MRC Biostatistics Unit
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Robinson College

July 2018

Declaration

This dissertation is the result of my own work between October 2014 and July 2018 and includes nothing which is the outcome of work done in collaboration except as specified in the text.

This thesis is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

Claire Louise Simons
July 2018

Acknowledgements

This work would not have been possible without the financial support of the MRC. I am especially grateful for the help, support, and encouragement of my supervisor, Dr Chris Jackson, whose patience and advice throughout the thesis has been invaluable. Discussions with my advisors, Professor Linda Sharples and Professor Simon Thompson, have enhanced this final thesis.

I would also like to thank Rajiv, Rosie, Maxine, Iryna, James, Bobby, Jacqui and Liz for their support, encouragement, and patience at various times throughout my thesis.

Finally, and most importantly, I would like to thank my parents, Kevin and Joan; my sister, Deborah; and my grandparents, John and Joan Thomson and Alan and Veronica Simons whose unconditional love and support has been with me throughout all the ups and downs of my life and my education.

Mum, Dad - this is it - I promise, no more degrees!

Abstract

Cost-effectiveness analysis (CEA) models, used to make health policy decisions, are usually subject to uncertainty. This thesis aims to develop statistical methods to quantify uncertainty and target where reducing uncertainty is most beneficial in a CEA. This enables policy decisions, based on the models, to be better informed. There is a focus on two areas: adherence to interventions and heterogeneity in treatment effects, which are often not modelled due to a lack of good data. A case study of treatment for patients with sleep apnoea is used to illustrate these methods and techniques.

Value of Information measures can help prioritise where to focus further research and estimate the expected benefits from a study of particular design and size. Until recently, it has been difficult to evaluate these quantities due to computational complexity. Various recently developed methods to calculate the expected value of information are summarised. Through an application to the case study, the importance of an adequate number of simulations to gain reliable results is highlighted.

Adherence to interventions is often neglected in CEAs due to limited and sparse data. Data on adherence to interventions for sleep apnoea is collected. Through Bayesian model-based meta-analyses, implemented by Markov Chain Monte Carlo simulation, the impact of modelling adherence to interventions on the CEA results is explored. Additionally, the value of collecting further information on adherence to interventions is calculated, indicating value in collecting data even at few time points, and in the early period of follow-up.

Another under explored area within CEAs is stratification of the optimal treatment decision. Here, the focus is on stratification based on continuous measures of disease severity, which may be associated with differential cost-effectiveness through variations in treatment effects. Aggregate and individual participant data on the impact of baseline covariates and treatment effects is summarised. Bayesian model-based meta-regression is used to explore stratification on one or two measure of treatment severity. The value of collecting further data on factors relating to stratification has been explored by using and extending recent non-parametric regression methods.

By using evidence synthesis methods, to make use of all available data, this thesis has found it is possible to incorporate uncertainty due to adherence to interventions and stratifications of treatment decisions into CEA models, allowing future research priorities to be assessed through value of information methods.

Table of contents

List of tables	xv
List of figures	xix
1 Introduction	1
1.1 Health economic theory	1
1.1.1 Economic evaluation	2
1.1.2 Probabilistic Sensitivity Analysis	9
1.1.3 Value of information quantities	11
1.2 Introduction to evidence synthesis	16
1.2.1 Introduction to meta-analysis	16
1.2.2 The general approach to Bayesian meta-analysis	18
1.2.3 Choice of priors for use in a Bayesian meta-analysis	19
1.3 Case study	20
1.3.1 Background to Sleep Apnoea	20
1.3.2 The Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea	22
1.3.3 The lifetime cost-effectiveness analysis	24
1.3.4 Results from the lifetime cost-effectiveness analysis	27
1.3.5 Recommendations for future work	29
1.4 Thesis aims and objectives	29
1.5 Structure of the thesis	30
2 Estimating value of information quantities	33
2.1 Introduction	33
2.2 The Expected Value of Perfect Information	36
2.2.1 Background and theory	36
2.2.2 Population-level Expected Value of Perfect Information	38

2.2.3	Monte Carlo error	40
2.3	The Expected Value of Perfect Partial Information	41
2.3.1	Background and theory	41
2.3.2	Methods of calculation	42
2.4	The Expected Value of Sample Information	55
2.4.1	Background and theory	55
2.4.2	Methods of calculation	57
2.5	The standard error of Expected Value of Perfect Partial Information estimates	65
2.5.1	The standard error for the Expected Value of Perfect Partial Information when estimated using regression based methods (Strong et al. (2014))	65
2.5.2	Extension of the Strong et al. (2014) standard error calculation to include Monte Carlo error	66
2.6	Exploring the population value of perfect information and their errors using the case study cost-effectiveness analysis	67
2.6.1	Using Monte Carlo error to calculate the optimal number of simulations for estimating the Expected Value of Perfect Information	67
2.6.2	The Expected Value of Perfect Partial Information for uncertain parameters in the case study	69
2.6.3	The population-level Expected Value of Perfect Information for the case study	70
2.7	Discussion	74
2.7.1	Limitations of the current methods	75
2.7.2	Future research priorities	77
2.7.3	Value of information in the rest of the thesis	79
2.8	Conclusion	79
3	Modelling adherence to interventions in a cost-effectiveness analysis	81
3.1	Introduction	81
3.1.1	Current guidance on including adherence to interventions in cost-effectiveness analyses	83
3.2	Literature review	84
3.2.1	Search criteria and methodology	85
3.2.2	Results	85
3.3	Methods for modelling adherence	89
3.3.1	A general time-to-event meta-analysis	89
3.3.2	Meta-analysis for modelling adherence to interventions	92

3.3.3	Goodness of fit measures	98
3.3.4	Methods for the inclusion of the results from a meta-analysis on adherence to a cost-effectiveness analysis	99
3.3.5	The value of collecting further information on adherence to interventions	100
3.4	Results	105
3.4.1	The time-to-event meta-analysis for adherence to MAD and CPAP as interventions for OSAHS	105
3.4.2	The cost-effectiveness analysis	106
3.4.3	The value of information	110
3.5	Discussion	113
3.5.1	Data limitations	115
3.5.2	Methodological issues	117
3.5.3	Future research priorities	117
3.6	Conclusion	119
4	Methods for stratifying the optimal treatment decision using non-binary measures of disease severity	121
4.1	Introduction	121
4.2	A literature review to guide modelling the impact of a baseline disease severity measure on the treatment effects with an intervention for patients with OSAHS	124
4.2.1	Search criteria and methodology	124
4.2.2	Results	125
4.3	Methods for estimating the impact of baseline measures on the treatment effect	129
4.3.1	Meta-regression theory	129
4.3.2	Network meta-regression models to estimate the impact of baseline measure of disease severity on the change in ESS with treatment for OSAHS	138
4.4	Application of the results of the meta-regression to the cost-effectiveness analysis	145
4.4.1	Exploring the relationship between the baseline ESS and the change in ESS between treatment arms analytically	146
4.5	Results of the network meta-regression and potential stratification of the optimal treatment decision	148
4.5.1	Potential stratification of the optimal intervention on one covariate: Baseline ESS	148

4.5.2	Potential stratification of the optimal treatment decision on two covariates: Baseline ESS and BMI	151
4.6	Discussion	154
4.6.1	Data limitations	157
4.6.2	Methodological issues	159
4.6.3	Future research priorities	161
4.7	Conclusion	161
5	The value of stratification and collecting further information to guide stratification	163
5.1	Introduction	163
5.2	The health economic value of stratification	165
5.2.1	Background and theory	165
5.2.2	An illustration of the impact of the population distribution and leakage on the case study	172
5.3	The value of collecting further information to guide stratified decision making	180
5.3.1	The Expected Value of Perfect Partial Information for parameters related to stratification	180
5.3.2	The Expected Value of Sample Information for parameters related to stratification	184
5.4	Discussion	196
5.4.1	Data limitations	197
5.4.2	Methodological issues	199
5.4.3	Future research priorities	200
5.5	Conclusion	201
6	Discussion and Conclusions	203
6.1	Overview of thesis findings	203
6.2	Challenges and limitations	206
6.3	Contributions to the literature	209
6.4	Future research priorities	210
6.5	Conclusion	212
	References	213
	Appendix A The Epworth Sleepiness Scale	233

Appendix B A cost-effectiveness model of treatments for patients with mild-moderate OSAHS	235
B.1 Background	235
B.2 Model structure	236
B.3 Inputs to the cost-effectiveness model	236
B.3.1 Baseline characteristics	236
B.3.2 Cardiovascular risk and sleep apnoea	239
B.3.3 Risk of a road traffic accident	241
B.3.4 Utilities	242
B.3.5 Adherence to interventions	242
B.3.6 Non-cardiovascular mortality rates	243
B.3.7 Costs	243
B.3.8 Treatment effects	247
B.4 Transition probabilities	248
B.4.1 Formulation of the transition probabilities	249
B.4.2 Utilities	253
B.4.3 Costs	254
B.5 Calculating the cost-effectiveness	255
Appendix C Search strategy for reviewing literature on adherence to MAD or CPAP as treatments for patients with OSAHS	257
Appendix D Reasons for inclusion and exclusion for papers reviewed for information on adherence to MAD or CPAP	259

List of tables

1.1	Types of Economic evaluation	3
1.2	Results from the case study cost-effectiveness analysis using 500,000 PSA samples calculated in R	28
2.1	Reviewed studies for the methods estimating the Expected Value of Perfect Partial Information along with their advantages and disadvantages	45
2.2	Reviewed studies for methods estimating the Expected Value of Sample Information along with their advantages and disadvantages	58
2.3	The Monte Carlo standard error and 95% interval for the expected loss for different sample sizes for the case study cost-effectiveness analysis	68
2.4	Monte Carlo standard error and 95% interval for the population expected loss for different numbers of PSA samples for the case study cost-effectiveness analysis	74
3.1	Definitions of adherence to interventions for patients with OSAHS in the studies identified by the literature review	87
3.2	Data summary for the proportion of participants adherent to interventions at measured times taken from included papers in the literature review	91
3.3	Posterior median and 95% credible intervals for the Random Effect mean pooled estimates along with the deviance statistics from the Bayesian time-to-event meta-analysis models using 100,000 Monte Carlo simulations for adherence to MAD and CPAP as interventions for OSAHS	109
3.4	Results of the implementation of adherence Model IV applied to the case study cost-effectiveness analysis using 100,000 PSA samples	111

4.1	The posterior median, 95% credible intervals and deviance statistics for the coefficients of the network meta-regression models on the impact of baseline ESS on the change in ESS using aggregate and individual participant data with 100,000 Monte Carlo simulations and pooled effects as the means from the posterior distributions	150
4.2	The posterior median, 95% credible intervals and deviance statistics for the coefficients of the network meta-regression models on the impact of baseline ESS and BMI on the change in ESS using aggregate and individual participant data with 100,000 Monte Carlo simulations and pooled effects as the means from the posterior distributions	155
5.1	Results of the exploration of the value of heterogeneity for three different population distributions for stratification on ESS using the results of the cost-effectiveness analysis presented in Chapter 4	176
5.2	Results of the exploration of the value of heterogeneity for six different population distributions for stratification on BMI and ESS using the results of the case study cost-effectiveness analysis in Chapter 4	179
B.1	Baseline characteristics for the cohort used in the case study cost-effectiveness analysis	237
B.2	Baseline model parameters for the case study cost-effectiveness analysis	237
B.3	Quantities calculated using the Framingham equations for use in the case study cost-effectiveness analysis	241
B.4	CHD and stroke parameters and their distributions used in the case study cost-effectiveness analysis	241
B.5	Underlying annual risk of road traffic accidents used in the case study cost-effectiveness analysis	242
B.6	Utility estimates used in the case study cost-effectiveness analysis	243
B.7	Costs associated with treatment by CM or MAD in 2011/2012 prices	245
B.8	Costs associated with initial use of CPAP in 2011/2012 prices	246
B.9	Costs associated with ongoing use of CPAP in 2011/2012 prices	246
B.10	Costs associated with coronary heart disease, stroke, and road traffic accidents in 2011/2012 prices	247
B.11	Modelled treatment effects and their distributions used in the case study cost-effectiveness analysis	248
B.12	Schematic table showing allowed transitions in the case study cost-effectiveness analysis	249

B.13	Notation used to derive the transition probabilities for the case study cost-effectiveness analysis	250
D.1	Papers for inclusion or exclusion in the review of adherence to MAD or CPAP as treatments for OSAHS after review of full text - as at January 2015	259

List of figures

1.1	A simplified schematic of the lifetime cost-effectiveness model structure developed by McDaid et al. (2009) to explore interventions for patients with OSAHS	25
1.2	The cost-effectiveness acceptability curve for patients with OSAHS treated with CM, MAD and CPAP for a range of thresholds for the case study cost-effectiveness analysis using 500,000 PSA samples	28
1.3	The cost-effectiveness acceptability frontier for patients with OSAHS treated with CM, MAD and CPAP for a range of thresholds for the case study cost-effectiveness analysis using 500,000 PSA samples	29
2.1	The Incremental Net Benefit for the MAD vs CPAP treatment comparison and QQ-plot for normality 100,000 PSA samples and a cost-effectiveness threshold of £20,000 per QALY gained	37
2.2	The expected loss and its 95% interval for different sample sizes in the case study cost-effectiveness analysis	68
2.3	The results of the EVPPI estimates for a selection of uncertain parameters in the case study cost-effectiveness analysis using 100,000 PSA samples	70
2.4	The results of the EVPPI estimates for parameters in the case study cost-effectiveness analysis relating to treatment effect for a range of PSA simulation sizes	71
2.5	The results of the standard error calculation for the EVPPI for various parameters relating to treatment effect in the case study cost-effectiveness analysis incorporating error due to the GAM and Monte Carlo error and presented for a range of PSA simulation sizes	71
2.6	Population-level EVPI values for a number of different parameters used compared to the baseline population EVPI value	73

3.1	Flow diagram of the selection process of the literature search for papers on adherence to MAD and CPAP as interventions for patients with OSAHS . . .	86
3.2	Observed adherence data on long-term adherence to MADs and CPAP as treatments for OSAHS taken from the studies analysed as part of the literature review	90
3.3	The prior distributions for the mean time to non-adherence and the doubling of the hazard rate for Model IV	96
3.4	Estimated five year posterior median and 95% credible interval adherence probabilities from four meta-analysis models modelling adherence to MAD and CPAP as treatments for OSAHS along with the pooled estimates	107
3.5	Observed adherence data and pooled results with the 75% credible intervals where relevant from four Bayesian time-to-event meta-analyses of adherence to MADs and CPAP as treatments for OSAHS using 100,000 Monte Carlo simulations, using a predictive study pooled estimate unless specified . . .	108
3.6	Population-level EVPPI quantities for the collection of data on adherence to MAD and CPAP as interventions for patients with OSAHS and other parameters using the case study cost-effectiveness analysis with adherence Model IV for ten years and the new predicted study pooled values from the results of the Bayesian meta-analysis	112
3.7	Population-level EVSI estimated quantities for the collection of data on adherence to MAD and CPAP as interventions for patients with OSAHS at single timepoints (Panel a) and at two timepoints (Panel b) using the case study cost-effectiveness analysis with adherence model IV for ten years and the new predicted study pooled values from the results of the Bayesian meta-analysis	114
4.1	Observed data on baseline ESS values and the difference in mean ESS between treatment arms extracted from the studies in the literature review .	127
4.2	Observed data on BMI values and the difference in mean ESS between treatment arms extracted from the studies in the literature review	128
4.3	Observed data on the individual patient baseline ESS values and the individual patient difference in ESS with treatment over time extracted from the studies where IPD was easily accessible	130
4.4	Observed data on the individual patient baseline BMI values and the individual patient difference in ESS with treatment over time extracted from the studies where IPD was easily accessible	131

4.5	Results of the case study cost-effectiveness analysis incorporating the results of the network meta-regression on the impact of baseline ESS on the change in ESS with treatment presented in terms of the INB between MAD and CPAP for all values of baseline ESS and a cost-effectiveness threshold value of £20,000 per QALY gained	152
4.6	Results of the case study cost-effectiveness analysis using the results of the network meta-regression presented as the probability each intervention is most cost-effective, for each plausible baseline ESS value and the intervention with the highest net monetary benefit at a cost-effectiveness threshold of £20,000 per QALY gained	153
4.7	Results of the case study cost-effectiveness analysis using the results of the network meta-regression presented in terms of the intervention with the highest net monetary benefit for all baseline ESS and plausible BMI combinations at a cost-effectiveness threshold of £20,000 per QALY gained	156
5.1	Three alternative population distributions of ESS used to calculate the value of stratification for the case study cost-effectiveness analysis	173
5.2	Two alternative population distributions for BMI used to calculate the value of stratification under a two variable stratification policy for the case study cost-effectiveness analysis	175
5.3	The weights defining the six population distributions of combined ESS and BMI values (assuming independence) used to calculate the value of stratification under a two covariate stratification policy for the case study cost-effectiveness analysis	177
5.4	The proportion of patients receiving their optimal treatment, as suggested by the case study cost-effectiveness analysis under two alternative leakage distributions compared to the base case	178
5.5	Point and 95% estimates of the Expected Value of Perfect Partial Information comparing the value of collecting further information on the treatment effect for different treatment comparisons for a population with uniform distribution(s) for the stratifier(s) of interest compared between a single and multiple calculation method	185
5.6	Results of the EVSI calculation for four study designs to collect further information to inform the four treatment comparisons compared between five baseline ESS distributions	193

5.7	Results of the EVSI calculation for a proposed cross-over study collecting aggregated data, four treatment comparisons, five baseline ESS distributions and three BMI distributions to collect further information to inform the relationship of the treatment effect to baseline ESS and BMI	194
5.8	Results of the EVSI calculation for a proposed cross-over collecting individual participant data, four treatment comparisons, five baseline ESS distributions and three BMI distributions to collect further information to inform the relationship of the treatment effect to baseline ESS and BMI . .	195
B.1	Simplified long term model structure developed by McDaid et al. (2009) to explore the life-time cost-effectiveness of treatments for OSAHS [134] . . .	238
B.2	The probability individuals are adherent to their intervention in the case study cost-effectiveness analysis based on data from Kohler et al. (2010)	244
B.3	The annual probability of a non-cardiovascular death used in the case study cost-effectiveness analysis for a male alive at age 50 using 2009-2011 interim life tables from the Office of National Statistics	245

Chapter 1

Introduction

Cost-effectiveness analysis is an integral part of how the National Institute for Health and Care Excellence (NICE), Scottish Medical Consortium (SMC) and All Wales Medical Strategy Group (AWMSG) in the United Kingdom (UK) produce their evidence based guidance and advice. This advice helps to ensure all National Health Service (NHS) patients in the UK have access to the most cost-effective treatments [147]. The ability to effectively allocate healthcare resources to maximise the health of the population, while keeping to a fixed budget, is essential to the success of the NHS. However, all decisions on resource allocation are subject to uncertainty.

This chapter introduces the research questions relating to uncertainty around resource allocation. It starts by outlining the key health economic and evidence synthesis concepts used throughout the thesis. It then introduces the case study, the cost-effectiveness of interventions for patients with sleep apnoea. Finally, it presents the objectives and structure of the rest of the thesis.

1.1 Health economic theory

Health economics can be seen as a combination of medical research, epidemiology, statistics and economics [12]. It applies economic theory of resource allocation, split into capital and labour, to health¹ and healthcare² [12]. This thesis explores some of the issues surrounding the optimisation of resource allocation to maximise population health in the presence of finite

¹As in Baio (2012) defined by the World Health Organisation (WHO) as "A state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity" [12, 254]

²Which is any aspect of healthcare given to an individual by a healthcare professional

resources to assist the payers in their decision making. This field of study is, specifically to health economics, known as economic evaluation [12, 24, 63, 84].

1.1.1 Economic evaluation

Economic evaluation is the comparison of two or more mutually exclusive alternative options in terms of their costs and consequences [63]. It aims to provide decision makers with the necessary information to make resource allocation decisions [63, 141]. Economic evaluation is used by the health systems of many countries including the UK, Australia and Canada [11, 33, 147]. In the UK, the NICE collate evidence on the clinical benefit and costs of health interventions³ [147].

As suggested by the definition above, the key components of an economic evaluation for a choice of alternative options are the costs and outcomes (consequences) [63]. The alternative options can be anything which use health resources - for example, diagnostic interventions, pharmaceuticals, surgical interventions, screening programmes, and public health interventions. In this thesis, the options are referred to as interventions, although theoretically this could refer to any of the above. As economic evaluation is a comparative procedure, all potential, new interventions need a comparator. Ideally this is every possible intervention for the patient. However, for practical reasons, in the UK, this is generally taken to be the current best practice in the NHS which, if there is no currently available intervention, could be no treatment or placebo [150].

The costs associated with each intervention depend on the perspective and context of the economic evaluation. These can include the cost of healthcare resources such as the interventions themselves, the time of medical professionals, and costs to the hospital, such as inpatient stay. Some economic evaluations also take into account the cost of the patients' and their family's time. This thesis takes the perspective required by NICE, that of the NHS and Personal Social Service (PSS) resources (i.e. excluding the cost of the time of patients and their families, out of pocket costs, lost productivity, and other public sector costs) [150]. However, NICE state that costs to government bodies, other than the NHS, or the cost of time provided by family members, friends, or partners can be included in exceptional circumstances [150].

³The NICE is an independent organisation responsible for providing national guidance. It was formed in 1999 and was established in Primary Legislation in 2013, meaning that although they operate independently from the government they are accountable to the Department of Health. The NICE uses evidence based guidance to provide the NHS (and those who use its services) advice on healthcare considered effective and good value [147].

Table 1.1 Types of economic evaluation

Economic Evaluation Method	Outcome Measured
Cost-consequence analysis (CCA)	Multiple outcomes reported disaggregated
Cost-minimisation analysis (CMA)	Equal effectiveness and safety of interventions
Cost-effectiveness analysis (CEA)	Natural units (such as life years, cases detected)
Cost-utility analysis (CUA)	Quality Adjusted Life Years (QALYs) (composite measure of mortality and morbidity)
Cost-benefit analysis (CBA)	Monetary valuation of outcomes

There are many types of economic evaluations. These differ in terms of how they treat outcomes. Outcomes measured in economic evaluations are typically the impact of the intervention(s) on patients' health. These can be measured in a number of ways. For example - life years, Quality Adjusted Life Years (QALYs) (a composite measure of mortality and morbidity allowing for consistent comparison between diseases), symptom free days, and disease specific measures [150, 157, 250]. Table 1.1, presents the main different types of economic evaluations and the outcome measure they use.

In a cost-consequence analysis the costs and outcomes (consequences) of the interventions are calculated and reported in a disaggregated way. It is the decision makers who then interpret the results in the way in which they wish. A cost-minimisation analysis, focuses on the cost differences between the interventions having assumed the health outcomes are identical for all interventions. A cost-minimisation analysis is not considered a full economic evaluations as it does not consider the joint distribution of cost and effects and there is rarely sufficient evidence to conclude that the treatments are equivalent in terms of outcomes [63]. A Cost Benefit Analysis (CBA) places a monetary values on health outcomes by using societies willingness to pay for a health benefit. Costs of new interventions are compared to the monetary value of the benefits and treatments. A new treatment is adopted if the benefits exceed the costs. In a Cost-Effectiveness Analysis (CEA) health outcomes are measured in natural units such as the primary outcome of a trial, life years, and cases detected. The costs of the interventions and the outcomes are used to calculate an Incremental Cost-Effectiveness Ratio (ICER) such as the cost per life year or cost per case averted. Finally, in a Cost Utility Analysis (CUA) bot mortality and morbidity outcomes are combined to create a single composite measure (QALYs). The use of QALYs mean the ICER, cost per QALY gained can be compared across different disease areas [12, 24, 84].

The terms CEA and CUA are used interchangeably in the literature and both support the adoption decisions made by NICE [24, 150]. The aim of a CEA is to find the intervention which minimizes the cost of generating a given level of health, or alternatively maximises the level of health for a specified budget, agreeing with the focus of NICE [24, 80]. Additionally, as stated in the NICE Methods Guidance (2013) CEAs and CUAs are more widely used than CBAs [150].

Decision Analytic Modelling

Data for making an optimal reimbursement decision can come from a variety of sources. NICE Methods Guidance (2013) outlines the current best practice in the UK for an analytic framework which is to use all ‘relevant’ available evidence [150]. These can include evidence from Randomised Controlled Trials (RCTs), observational studies, and expert opinion. However, different sources of data are seen to be preferential in terms of the risk of bias and internal validity. This is known as the ‘hierarchy of evidence’, of which there are many forms [68, 85, 144, 186]. However, they all take a similar format: weaker studies at the bottom such as basic science and case reports, followed by case-control studies, and then RCTs with systematic reviews and meta-analyses seen as the data sources with the least risk of bias in best internal validity [144].

Mathematical methods are typically needed to synthesise the data [63]. This mathematical framework is known as Decision Analytic Modelling (DAM) [24]. Section 1.2 introduces meta-analysis methods which is one method that can be used to combine data of the same kind from multiple sources for inclusion in an economic model and is used later in the thesis (Chapters 3 and 4).

A DAM is a function that links data synthesised from many sources and its associated uncertainty to the outcomes of interest, for example costs and QALYs linking parameter uncertainty to decision uncertainty [239]. The results presented from DAMs are conditional on the input data and the assumptions underpinning the model [239].

The decision problem in a Bayesian decision theoretic framework is outlined in Baio and Dawid (2011) and Baio (2013) [12, 13]. Assume a set of interventions $j \in \mathcal{J}$, where the interventions are mutually exclusive. For each intervention j there are individual and population-level responses, y_j and $Y_j = E(y_j)$ respectively. These can be uni- or multivariate. Typically, in CEAs the response (y_j) is a bivariate measure of a clinical outcome (e_j) and a

cost (c_j).

In a Bayesian decision theoretic CEA the decision maker wants to decide which j to give a population on the basis of the outcomes Y_j .

Let $\boldsymbol{\theta} = (\theta_0, \dots, \theta_S)$ be the set of S population parameter quantities that govern the outcomes, i.e. included in the DAM. $\boldsymbol{\theta}$ is given a distribution which represents the current uncertainty in the model by reflecting all currently available data \mathcal{D} . See Section 1.2 for further information on how values of $\boldsymbol{\theta}$ are derived and how Y_j is defined in terms of $\boldsymbol{\theta}$. However, briefly it is unlikely a single study can provide information on $\boldsymbol{\theta}$. Therefore, statistical methods that take into account all available data, \mathcal{D} are often used. The information for each element of $\boldsymbol{\theta}$ can then be combined to define Y_j through the DAM. The joint distribution of $\boldsymbol{\theta}$ is $p(\boldsymbol{\theta}|\mathcal{D})$ which can be thought of as a prior on $\boldsymbol{\theta}$ or as the posterior distribution of $\boldsymbol{\theta}$ given \mathcal{D} . The distributions of $\boldsymbol{\theta}$ are indirectly relevant to the decision process through the probability distribution $p(y_j|\boldsymbol{\theta})$ which represents the variability in the individual-level response and the implied population-level outcome $Y_j = E(y_j)$.

Typically, the individual-level outcome y_j for intervention j , is defined in terms of effects (e) (QALYs) and costs (c), so for each j , $y_j = (e_j, c_j)$. The aim of the CEA is to compare the interventions on the basis of their outcomes at a population-level. The incremental mean effectiveness and cost of intervention 1 compared to intervention 0 over individuals, when $j = 0, 1$ (although this can be easily extended to the case with more than two interventions) are:

$$\begin{aligned}\Delta_E &:= E[e_1] - E[e_0] \\ &= \bar{e}_1 - \bar{e}_0 \\ \Delta_C &:= E[c_1] - E[c_0] \\ &= \bar{c}_1 - \bar{c}_0\end{aligned}$$

This thesis does not cover CEAs based on models for individual-level cost-effectiveness outcomes $p(y_j|\boldsymbol{\theta})$. These are typically carried out using trial data whereas the focus of this work is on cohort model based CEAs. In a cohort model, a DAM using all available data, expressed by the population parameters $\boldsymbol{\theta}$, is used to generate the population level outcomes, i.e. $Y_j = h_j(\boldsymbol{\theta})$ where $h_j(\boldsymbol{\theta})$ is a DAM making use of $\boldsymbol{\theta}$.

Δ_E and Δ_C are functions of (unknown) θ so are random variables under a Bayesian framework. Taking a further expectation with respect to θ gives the posterior mean incremental effects and costs between interventions 1 and 0 for the population:

$$\begin{aligned} E_{\theta} [\Delta_E] &= E_{\theta} [\bar{e}_1 - \bar{e}_0] \\ E_{\theta} [\Delta_C] &= E_{\theta} [\bar{c}_1 - \bar{c}_0] \end{aligned}$$

The results from a CEA can be presented as an ICER, which under a Bayesian framework is:

$$ICER = \frac{E_{\theta} [\Delta_C]}{E_{\theta} [\Delta_E]}$$

An intervention is said to be cost-effective if $ICER < \lambda$, where λ is the cost-effectiveness threshold (£ per QALY gained). This requires an explicit value for the maximum the decision maker is willing to pay for an extra unit of outcome, λ . In England and Wales, NICE Methods Guidance states that explicit reasons⁴ should be given if an intervention that is cost-effective at a threshold of less than £20,000 per QALY gained is not accepted [150]. Interventions that are cost-effective at thresholds between £20-30,000 per QALY gained need to be considered further and have an increasing degree of certainty around the cost-effectiveness. Any intervention accepted for use on the NHS above a threshold of £30,000 per QALY gained needs an increasingly strong argument that it is an effective use of resources using the factors above. Hence in practice $\lambda = £20,000 - 30,000$ per QALY gained is used⁵ [150, 151].

Although the ICER has been widely used historically it has some major limitations [205]. For example, if an intervention is more effective and less costly (e.g. $(E_{\theta}[\Delta_E], E_{\theta}[\Delta_C]) = (-1, +100)$) or more effective and less costly (e.g. $(E_{\theta}[\Delta_E], E_{\theta}[\Delta_C]) = (+1, -100)$), the ICER has the same sign and value [12]. Obviously, one situation is more desirable and it is not easy to identify the optimal treatment by observing the ICER. If both the incremental costs and the incremental QALYs are negative (i.e. the new intervention is cheaper and reduces QALYs) then it is possible for the resulting ICER to be under the threshold λ . However, deeming such an intervention as cost-effective raises ethical questions around

⁴These include explanations on the presence of any reasons indicating change in Quality of Life inadequately represents the health gain; whether the intervention is innovative; and whether the intervention gives additional substantial benefits not featured in the health gain.

⁵A report by Claxton et al. (2015) estimate the 'central' empirical threshold was £12,936 per QALY gained by looking at routinely collected data to estimate the relationship between NHS expenditure and changes in quality of life [41]. They also found the probability the threshold was less than £20,000 per QALY gained was 0.89, indicating the threshold currently used in practice is too high

whether people should lose health to save money [83]. Additionally, interval estimates are difficult to interpret.

Instead, a Bayesian decision theoretic framework is preferred where the optimal intervention is the j which maximises the expected utility over the population. Assuming $\mathcal{J} = (0, 1)$ and $u(Y_j)$ is the utility value of giving intervention j and getting response Y from the population, then the expected utility under current information \mathcal{D} for a population given intervention j is:

$$\mathcal{U}^j := E_{\boldsymbol{\theta}} [u(Y_j) | \mathcal{D}]$$

The optimal utility is the maximum of the expected utilities, \mathcal{U}^j , across all j :

$$\mathcal{U}^* = \max_j \mathcal{U}^j$$

The Expected Incremental Benefit (EIB), the additional expected utility gained or lost by giving the population $j = 1$, compared to $j = 0$, can be defined as [12]:

$$\begin{aligned} EIB &:= \mathcal{U}^1 - \mathcal{U}^0 \\ \mathcal{U}^* &= \max \{EIB, 0\} + \mathcal{U}^0 \end{aligned}$$

When $EIB > 0$, intervention 1 is optimal in terms of maximising expected utility. In practice, a specific form is required for the utility function. A commonly used utility function, assuming the decision maker is risk neutral, for an individual is [12, 24, 209]:

$$u(y_j) = \lambda e_j - c_j$$

and for the population:

$$\begin{aligned} u(Y_j) &= E[u(y_j)] \\ &= \lambda \bar{e}_j - \bar{c}_j \end{aligned} \tag{1.1}$$

As $\boldsymbol{\theta}$ is a random variable under a Bayesian framework, expectations with respect to $\boldsymbol{\theta}$ need to be taken:

$$E_{\boldsymbol{\theta}} [u(Y_j)] = \lambda E_{\boldsymbol{\theta}} [\bar{e}_j] - E_{\boldsymbol{\theta}} [\bar{c}_j]$$

where λ is an explicit value for the threshold willingness to pay. This utility function (Equation 1.1) is commonly called the Net Monetary Benefit (NMB) of intervention j on the

population⁶ [12, 24, 209]. Going forward, $u(Y_j)$ shall be denoted $NB(j, \theta)$. The Incremental Net Benefit (INB) for the population treated with $j = 1$ compared to $j = 0$ is:

$$\begin{aligned} INB(\theta) &= NB(j = 1, \theta) - NB(j = 0, \theta) \\ &= [\lambda \bar{e}_1 - \bar{c}_1] - [\lambda \bar{e}_0 - \bar{c}_0] \\ &= \lambda [\bar{e}_1 - \bar{e}_0] - [\bar{c}_1 - \bar{c}_0] \\ &= \lambda \Delta_E - \Delta_C \end{aligned}$$

The focus, in terms of the analysis, is the expected INB, the EIB which is defined as [12]:

$$\begin{aligned} EIB &= E_{\theta} [NB(j = 1, \theta)] - E_{\theta} [NB(j = 0, \theta)] \\ &= E_{\theta} [\lambda \Delta_e - \Delta_c] \\ &= \lambda E_{\theta} [\Delta_e] - E_{\theta} [\Delta_c] \\ &= E_{\theta} [INB(\theta)] \end{aligned}$$

If $EIB > 0$, $j = 1$ is cost-effective relative to $j = 0$. The ICER is the λ such that $EIB = 0$, i.e. the λ where the decision maker is indifferent between interventions.

As with the NHS, if the aim of the healthcare system is to maximise the health of the population subject to a finite budget, treatment decisions should be based on expectations [145]. Claxton (1999) states if traditional ‘statistical significance’ rules were used for making decisions on cost-effectiveness then opportunity costs would be created [37]. The paper goes on to say patients have to be given one of the treatments so a decision has to be made, and it is a ‘historical accident’ which treatment is considered current best practice. Therefore, to maximise the population health on average, the intervention, j , with the greatest expected maximum NMB, that is:

$$\arg \max_j E_{\theta} [NB(j, \theta)]$$

should be the most cost-effective intervention. However, implementation on this basis has risks. Therefore, Claxton (1999) recommend consideration of whether further data, \mathcal{E} , could

⁶This is the case under an extra-welfarist approach, where medical care should be compared against other types of healthcare. This can be compared to a welfarist approach whether medical care should be judged against all other goods. There is a range of literature on the most appropriate approach to take not limited to Brouwer (2008), Gyrd-Hansen (2005) and Buchanan (2015) [29, 30, 88]

be collected to help reduce decision uncertainty [37]. This is discussed further in Section 1.1.3.

1.1.2 Probabilistic Sensitivity Analysis

As collecting further information can delay implementation and incur costs, a sensitivity analysis should be carried out on the CEA to motivate the optimal decision as well as describing the uncertainty around it [12, 24]. There are two main types of sensitivity analysis: deterministic and probabilistic [12, 84].

A deterministic sensitivity analysis involves changing one or more of the parameters in the CEA to pre-defined values and recalculating the expected utility [24, 84]. This can be difficult if θ is large and/or the model is complex. Deterministic sensitivity analysis does not take into account any correlation between elements of θ and relies on the modeller to set the parameter values to test the robustness of the results [24, 84]. Deterministic sensitivity analyses cannot characterise the nature of the uncertainty. They can also be very time consuming should multiple different scenarios of θ need to be considered.

A more appropriate technique, often called Probabilistic Sensitivity Analysis (PSA), is equivalent to using the Bayesian Decision Theoretic framework (Section 1.1.1) [12, 24, 84]. The parameters θ are random quantities given distributions reflecting current evidence, \mathcal{D} . $INB(\theta)$ and $NB(j, \theta)$ are random quantities whose probability distributions are entirely dependent on $p(\theta|\mathcal{D})$ [12].

A PSA is typically implemented in practice by using Monte Carlo simulation methods. Firstly, $\theta^{(k)}$, $k = 1, \dots, K$ is sampled from $p(\theta|\mathcal{D})$. The CEA is run assuming $\theta^{(k)}$ from $p(\theta|\mathcal{D})$ is the realised value of θ for that simulation. The outcomes, in this case the cost and effect pairs, are recorded for each simulation. These can be used to calculate $NB(j, \theta)$ or $INB(\theta)$ (Section 1.1.1) [12, 24, 84].

Note that, as $INB(\theta)$ does not typically take any particular distributional form due to the structure of the DAM, the expectation over θ of $INB(\theta)$, the EIB, does not equal the INB of the expectation of the values of θ [24, 179]:

$$E_{\theta}(INB(\theta)) \neq INB(E(\theta))$$

The optimal treatment is chosen as the j with the maximum $E_{\boldsymbol{\theta}}(NB(j, \boldsymbol{\theta}))$. By taking into account the distributions and uncertainty within $\boldsymbol{\theta}$ rather than just using point estimates of $\boldsymbol{\theta}$ a PSA helps to improve decision making.

Presenting results from a PSA

Results from a PSA are often presented using a Cost-Effectiveness Acceptability Curve (CEAC) [12, 71, 72, 232]. A CEAC presents the probability an intervention is the most cost-effective for a range of thresholds (λ), i.e. for each λ the probability each intervention has the highest NMB. That is:

$$\begin{aligned} CEAC(\lambda) &= Pr(INB(\boldsymbol{\theta}) > 0) \\ &= Pr(\lambda \Delta_e - \Delta_c > 0) \end{aligned}$$

$\lambda \Delta_e - \Delta_c > 0$ indicates $j = 1$ is optimal, thus the CEAC presents the probability that knowing $\boldsymbol{\theta}$ with certainty would not change the optimal treatment decision. CEACs provide a clear visual representation of the uncertainty around the most cost-effective intervention at various thresholds and show the λ where the intervention with the highest probability of being cost-effective changes. As $\lambda \rightarrow 0$, $CEAC(\lambda)$ tends to the probability $j = 1$ is cheaper than $j = 0$. Conversely, as $\lambda \rightarrow \infty$ the CEAC tends to the probability $j = 1$ is more effective than $j = 0$ [12, 24, 71, 72].

The CEAC described above is for $j = \{0, 1\}$. However, there can be more than two interventions. Presentation of multiple CEACs is the same conceptually as with two interventions. However, the CEAC shows the probability each intervention, j , is the most cost-effective:

$$CEAC_j(\lambda) = P\left(NB(j, \boldsymbol{\theta}) = \max_j NB(j, \boldsymbol{\theta})\right)$$

A CEAC is illustrated in Figure 1.2 (Page 29) for three interventions. At values of $\lambda \approx \text{£}5,000 - 15,000$ per QALY gained there is the most uncertainty around which intervention is optimal. However, for very low values of λ this uncertainty is lower and further information on $\boldsymbol{\theta}$ is unlikely to alter the cost-effectiveness decision.

A major drawback to CEACs is they only address how likely it is that resolving uncertainty in $\boldsymbol{\theta}$ will change the optimal treatment decision [71]. The NMBs for each intervention are not included, thus CEACs do not indicate which treatment is optimal for each λ . The Cost Effec-

tiveness Acceptability Frontier (CEAF) presents the probability the optimal intervention j , chosen under the rule of maximising the expected NMB for each λ , is the most cost-effective [71, 72]. If the NMB has a skewed distribution the optimal intervention in terms of NMB may not necessarily correspond to the intervention with the highest probability of being cost-effective [24].

Figure 1.3 (Page 29) illustrates a CEAF for three interventions. The presence of discontinuities at $\lambda \approx £7,500$ per QALY gained indicates the optimal intervention in terms of the NMBs is not necessarily the same as the intervention with the highest probability of being cost-effective. This highlights the difference between a CEAC and a CEAF.

1.1.3 Value of information quantities

When considering the results of a CEA, two decisions need to be made - (i) should the intervention be implemented given existing evidence? and (ii) is further information required to support the implementation decision in the future? [24, 165]. Sections 1.1.1 and 1.1.2 answer the first question. Here, the second is considered.

If a suboptimal decision was made as to whether an intervention should be implemented in the NHS costs would be incurred in terms of the resource cost and the health benefits foregone by patients [24, 37]. The expected cost of uncertainty, and equivalently the expected opportunity loss, are related to the probability of making a wrong decision and its consequences in both health and money. This can be calculated and interpreted as the Expected Value of Perfect Information (EVPI) - perfect information on all parameters would eliminate the probability of making a wrong decision [12, 24, 37, 40].

Bayesian decision theory and value of information analysis can be used to create an analytic framework to answer the questions of interest. While value of information has only been implemented in a health economics context since the 1990s, it was mentioned in statistical decision theory in the 1950s [37–39, 177]. This section gives details on various quantities relating to the collection of further information - EVPI, Expected Value of Perfect Partial Information (EVPPI) and Expected Value of Sample Information (EVSI), laying the foundations for Chapter 2.

Expected Value of Perfect Information (EVPI)

Under current information, treatment decisions need to be made before uncertainties are resolved so the decision made is the best one, on average, with uncertain parameters. However, if perfect information was available on all parameters, a correct decision is guaranteed. The EVPI is the difference between the expected NMB under perfect and current information [2, 24, 70, 171, 194, 226].

Formally, as explained in Briggs et al. (2006), let j index the interventions, with S uncertain parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_S)$ in the CEA [24]. Let $NB(j, \boldsymbol{\theta})$ be the NMB for a threshold λ with intervention j and uncertain parameters $\boldsymbol{\theta}$. Under existing information, the optimal intervention is the j generating the maximum expected NMB:

$$\max_j E_{\boldsymbol{\theta}} [NB(j, \boldsymbol{\theta})]$$

Under perfect information, $\boldsymbol{\theta}$ is known a priori, so the j maximising the ‘true’ NMB can be selected:

$$\max_j NB(j, \boldsymbol{\theta})$$

However, in practice, the true value of $\boldsymbol{\theta}$ is unknown so the average of the maximum NMB over the joint distribution of $\boldsymbol{\theta}$ needs to be taken:

$$E_{\boldsymbol{\theta}} \left[\max_j NB(j, \boldsymbol{\theta}) \right]$$

The EVPI for an individual patient’s treatment decision is the difference between the expected value of the decision made with perfect information on $\boldsymbol{\theta}$ and the treatment decision made using existing information:

$$EVPI = E_{\boldsymbol{\theta}} \left[\max_j NB(j, \boldsymbol{\theta}) \right] - \max_j E_{\boldsymbol{\theta}} [NB(j, \boldsymbol{\theta})] \quad (1.2)$$

The EVPI is the expected cost of uncertainty. For a healthcare system that aims to maximise the population health subject to a budget constraint, as with the NHS⁷, the EVPI is the maximum the healthcare system should pay for perfect information on the treatment decision in the future (per person) [24, 145].

⁷The NHS also aims to provide everyone who needs it free healthcare, i.e. it also cares about the distribution of health

Population-level Expected Value of Perfect Information

The EVPI (Equation 1.2) is the value each time a treatment decision needs to be made, i.e. for each patient or each patient episode [24]. Any future research which creates additional evidence to inform the treatment decision can be used to benefit all current (in the case of chronic diseases) and future patients. Therefore, it is important to consider the EVPI for the population who would benefit from further research [24].

Calculation of population-level EVPI requires additional information - the expected future lifetime of the technology (T); estimates of the incidence of the disease over T (I_t); the current prevalence of the disease (I_0); the population at risk of the disease (P); and a discount rate (i) - typically 3.5% in England and Wales [150]. These values depend upon the nature of the disease. Whether the incident and/or prevalent population are included in the population-level EVPI depends on the nature of the disease (acute or chronic) and whether the patients in the trial are able to benefit from the information collected at a later date. The population-level EVPI (popEVPI), assuming I_t is available over discrete time units t , is:

$$popEVPI = EVPI \times P \times \left(I_0 + \sum_{t=1, \dots, T} \frac{I_t}{(1+i)^t} \right) \quad (1.3)$$

The population-level EVPI can be directly compared to the cost of research. It is a necessary, but not sufficient, condition for further research that the costs of research is less than the population EVPI.

There can be difficulties in estimating the values required in the calculation of the population-level EVPI. Estimating the appropriate time horizon for the population-level EVPI is challenging. Philips et al. (2008) found finite time horizons for decision problems are often used as a proxy for the complex and uncertain future technology change [165]. These issues and more background on population-level EVPI are presented in Chapter 2.

Expected Value of Perfect Partial Information (EVPPI)

It is infeasible to resolve all uncertainty on all elements of θ in the CEA model [12, 24, 39]. Therefore, the EVPI is regarded as a theoretical quantity, although it can still be useful as an upper bound for the value of future research. It may be of more interest to look at the value of reducing or eliminating uncertainty for a subset of uncertain parameters. The value of resolving uncertainty on a subset of parameters is known as the EVPPI (the expected value of perfect partial information or the expected value of perfect information for parameters). The

EVPI can help focus future research on the most important types of evidence by identifying those (sets of) parameters where more precise estimates are most valuable. However, as for EVPI, eliminating the uncertainty in the parameters requires an infinite sample size.

Similar to the EVPI, the EVPPI is the difference between the expected value of the decision with perfect and current information on the parameter(s) of interest [2, 171, 226]. Formally, we are interested in the value of perfect information for a parameter or a subset of all uncertain parameters (ϕ) with θ the set of all uncertain parameters and $\phi \subseteq \theta$ [24]. Under perfect information, the true value of ϕ is known, so the expected NMB of the decision is the j with the maximum expected NMB averaged over the remaining uncertain parameters in the model, $\bar{\phi}$ ($\bar{\phi} \cup \phi = \theta$). Letting $NB(j, \phi, \bar{\phi})$ be the NMB of intervention j with uncertain parameters ϕ and $\bar{\phi}$:

$$\max_j E_{\bar{\phi}|\phi} [NB(j, \phi, \bar{\phi})]$$

Given perfect information on ϕ , we want the expected NMB over the remaining parameters $\bar{\phi}$, hence the expectation over $\bar{\phi}|\phi$. The true values of ϕ are not generally known in practice, so the expected value of a decision under perfect information is found by averaging the maximum expected NMB over ϕ :

$$E_{\phi} \left[\max_j E_{\bar{\phi}|\phi} [NB(j, \phi, \bar{\phi})] \right]$$

As $\phi \cup \bar{\phi} = \theta$, the expected value under current information is the same as for EVPI, the optimal decision is the choice of j generating the maximum expected NMB:

$$\max_j E_{\theta} [NB(j, \theta)]$$

The EVPPI for the parameter(s) ϕ is the difference between the expected value of the decision made with perfect information about ϕ and the decision made on the basis of existing information:

$$EVPPI_{\phi} = E_{\phi} \left[\max_j E_{\bar{\phi}|\phi} [NB(j, \phi, \bar{\phi})] \right] - \max_j E_{\theta} [NB(j, \theta)] \quad (1.4)$$

Expected Value of Sample Information (EVSI)

The EVPPI is an upper bound for the value of future research on a set of parameters, ϕ . Eliminating uncertainty in ϕ is unlikely unless an infinite sample size is available. However,

it may be possible to reduce uncertainty in ϕ by carrying out a study with a particular design and set size m [2, 24, 216]. The EVSI is the expected value of conducting a particular research study design to reduce uncertainty in ϕ using a sample size of m . The expected benefits of the study can be compared to the expected costs of undertaking the study. If the expected benefits exceed the expected costs of research there is value in carrying out the study [24]. The EVSI and costs of research can be calculated for a number of different study designs and sample sizes enabling estimation of the optimal study [2, 24, 216].

The EVSI is the expected difference between the value of the optimal decision based on some sample of data informative for a subset of inputs ϕ , and the value of the decision made with current information [2, 171]. Formally, as outlined in Strong et al. (2015), there are j interventions and a set of uncertain parameters θ [216]. Let \mathbf{X} be the uncollected data generated from the proposed study as a vector of random variables assumed to arise from a statistical model with parameter(s) of interest $\phi \subseteq \theta$. An observation of \mathbf{X} can be used to learn about ϕ .

Under current information, the optimal decision is the choice of j generating the maximum expected NMB:

$$\max_j E_{\theta} [NB(j, \theta)]$$

If additional data \mathbf{X} are collected the optimal treatment is the one with the greatest NMB averaged over the joint posterior distribution of θ , conditional on the data $\theta|\mathbf{X}$:

$$\max_j E_{\theta|\mathbf{X}} [NB(j, \theta)]$$

Before carrying out the proposed study \mathbf{X} is uncollected, so unknown, therefore an average over all possible datasets arising from the study needs to be taken:

$$E_{\mathbf{X}} \left[\max_j E_{\theta|\mathbf{X}} [NB(j, \theta)] \right]$$

The EVSI is the difference between the expected NMB having collected the additional data and the expected NMB under current information:

$$EVSI = E_{\mathbf{X}} \left[\max_j E_{\theta|\mathbf{X}} [NB(j, \theta)] \right] - \max_j E_{\theta} [NB(j, \theta)] \quad (1.5)$$

The EVSI calculates can be used in an optimisation exercise to find the optimal sample size by maximising the Expected Net Benefit of Sampling (ENBS) over m [2]:

$$ENBS_X(m) = popEVSIX(m) - Cost(m)$$

where $popEVSIX$ is the population level EVSI for a study of size m and $Cost(m)$ is the cost (in money, health foregone and opportunity costs) of undertake the proposed study of size m . The aim is to find the m that maximises $ENBS_X(m)$.

1.2 Introduction to evidence synthesis

It is unlikely a single piece of evidence (for example, a RCT or an observational study) will be sufficient to estimate values (for example: treatment effect, disease incidence and progression, mortality, Health Related Quality of Life (HRQoL), and costs) to be used in a CEA [57]. Therefore, it is useful to have methods that can combine data to make inferences. Evidence synthesis is the broad term used to describe any inference made on a quantity which uses more than one data source simultaneously [126, 220].

The multiple data sources used in evidence synthesis are generally collected through a systematic literature review. The resulting data from the review can be summarised qualitatively or quantitatively. Meta-analysis and meta-regression models are often used as a specific statistical methodology to combine estimates of quantities of the same kind from multiple studies [126]. This is in contrast to decision models (Section 1.1.1) that can combine different types of data.

Throughout this thesis meta-analysis and meta-regression models will be used to synthesise various quantities. The aim of this section is to outline the basic premise of meta-analysis particularly applied to time-to-event outcomes and continuous treatment effects. Meta-analysis models are applied to observational data on time to treatment non-adherence in Chapter 3. Meta-regression models for continuous treatment differences where the available data is a combination of Aggregate Data (AD) and Individual Participant Data (IPD) from RCTs are used in Chapter 4.

1.2.1 Introduction to meta-analysis

Meta-analysis is a statistical approach to combining the results of multiple data sources. It aims to obtain a summary estimate of the quantity of interest using all available data [81, 98, 203, 205]. The term ‘effect’ is commonly used in meta-analysis literature to describe the quantity being pooled. However, treatment effects are not always the quantity of interest,

for example in Chapter 3 the pooled quantity is the time to non-adherence. This chapter uses the term effect to represent the quantity being pooled.

It is possible to undertake meta-analysis under both Bayesian and classical approaches, however work in this chapter and this thesis use the Bayesian approach.

In a classical approach the meta-analysis produces a point estimate with an associated 95% confidence interval by weighting the estimates from studies using various approaches with the weights related to the study size [81, 96, 98, 205]. The Bayesian approach combines prior beliefs about the pooled effect with information from the studies being synthesised to obtain the posterior distribution of the pooled effect where the precision of this estimate should depend on the study size, often the standard error of the treatment effect [81, 98, 110, 203, 205].

The advantages of a Bayesian approach include being able to easily present inferences which fully take into account uncertainty about all unknown quantities, such as the extent of between study heterogeneity. Additionally, there is readily available software such as JAGS and BUGS which can be used to carry out Markov Chain Monte Carlo (MCMC) [126, 169]. Software is also available for a classical meta-analysis, such as the *mvmeta* package in STATA and Review Manager [224, 244]. The posterior probabilities produced by a Bayesian analysis have also been argued to be easier to interpret than the p-values from the classical approach [205].

In setting a distribution representing prior beliefs researchers are required to consider what they would expect the plausible outcomes from the meta-analysis to be which can be challenging [81, 205, 220]. For example, if the outcome is a treatment effect that is restricted by a scale (e.g. can only take values between 0 and 10) the prior should reflect this. Elicitation of priors is non-trivial and different subjective priors can lead to different inferences [81, 205, 220]. However, sufficient data can overcome the influence of the prior. See Section 1.2.3 for more information on choosing priors for use in a Bayesian meta-analysis.

In both Bayesian and classical meta-analysis there are two main approaches: fixed effects and random effects [220]. Under a fixed effects model, each study is used to estimate an effect that is assumed to be common between studies [98, 220]. The differences between the data from each study are assumed to be due to sampling error. In the classical approach this is often calculated using an inverse-variance-weighted method [43]. The pooled effect is estimated as a weighted average of the study effects, with the weight inversely proportional

to the variance of the study specific estimate [110].

When a random effects approach is taken, differences in the data between studies are assumed to be due to both sampling error and heterogeneity between studies [98, 203]. Heterogeneity is due to studies having differences, such as in their population, the interventions, and their outcomes. This means it is not viable to assume a common parameter for all studies but instead it is assumed these parameters come from the same distribution [98]. Quantities representing this underlying heterogeneity are known as random effects [98, 203]. The pooled effect is still a weighted average of the study effects, either implicitly in a Bayesian approach or explicitly in a classical approach [98, 203, 205]. However, the weights take into account both the uncertainty in the estimates from each study and the random effects variance [98, 110, 203, 205].

1.2.2 The general approach to Bayesian meta-analysis

More formally, suppose the data from study i are generated from a model with parameters θ_i . We want to obtain a pooled estimate of the quantity described by θ_i using data from all studies. Exchangeability is a key assumption made in Bayesian random effects meta-analysis. Let the set of unit specific parameters be θ_i where $i = 1, \dots, N$ indexes the set of studies. Under exchangeability, it is assumed the θ_i 's arise from a common population distribution, with unknown parameters and appropriate priors, i.e. the θ_i 's are similar but not identical [81, 126, 205]. As an example, for a parameter θ_i assume:

$$\theta_i \sim N(\mu, \sigma^2)$$

where μ and σ^2 are given prior distributions and the data from study i , y_i , are generated conditionally on θ_i . Under the assumption of exchangeability, information about θ_i can be learned through the direct information, y_i , and the indirect information consisting of y_j where $j \neq i$ which inform the population distribution parameters $\phi = (\mu, \sigma^2)$ [81, 126, 205]. The θ_i 's are the random effects with σ^2 expressing the extent of heterogeneity.

The joint prior for all unknown parameters, $p(\theta_1, \dots, \theta_N, \phi)$, takes into account the prior distribution for the population parameters and the exchangeability assumption for each unit specific parameter [12, 81, 98, 205, 220]. That is:

$$p(\theta_1, \dots, \theta_N, \phi) = p(\phi) \prod_{i=1}^N p(\theta_i | \phi)$$

The model is defined in layers, explaining why it is often referred to as an hierarchical model [126]. The pooled estimates from a meta-analysis can be defined in a number of different ways [81, 98, 203, 243]. Different types of pooled estimates are described in more detail in Chapter 3.

If there are more than two treatments that need to be compared, a network meta-analysis can be undertaken [56, 110, 227]. A network meta-analysis can find pooled effects for all treatment comparisons including those there is no direct data on and can include studies with more than two arms [56, 110, 227]. Further, the pooled effects for a particular treatment comparison can borrow strength from other treatment comparisons in the network [56, 110, 227]. Further information on network meta-analysis and multi-arm studies is presented in Chapter 4.

1.2.3 Choice of priors for use in a Bayesian meta-analysis

Choosing the prior distributions for the meta-analysis model is an important stage of the process. The prior should reflect the range of plausible values for the parameters. O'Hagan (2006) and Gelman et al. (2014) advocate the use of weakly informative priors [81, 158]. A weakly informative prior is one that can "provide approximations to a more meticulous Bayesian analysis" [158]. If the data are relatively strong in comparison to prior information then a weakly informative prior should give essentially the same posterior distribution as a more informative prior. Gelman et al. (2014) define a weakly informative prior to be one which provides information sufficient to ensure the posterior distribution makes sense but is intentionally weaker than the available knowledge [81]. These could be intentionally weaker than the data for ease of calculation. They could be feasible approximations to the results of a more formal analysis for the priors. However, sensitivity to weakly informative priors should be checked. It is often the case that prior knowledge is much weaker than the available data and so as long as the prior distribution is plausible the data should override the prior [81]. Conducting a formal elicitation to gain an informative prior is often considered not cost-effective due to the time and effort required in its estimation [81]. However, this is an active area of research.

It is well-known the results of meta-analyses can be sensitive to chosen priors [81]. Therefore, ideally, a number of different weakly informative priors should be considered to identify the impact the choice of priors has on the data. This was done in the preliminary analysis for the work in Chapters 3 and 4.

1.3 Case study

This thesis makes extensive use of a case study CEA on treatment for patients with sleep apnoea. This section outlines the disease, its treatments, the base case CEA structure, and results.

1.3.1 Background to Sleep Apnoea

Obstructive Sleep Apnoea-Hypopnoea (OSAH) is defined by the American Academy of Sleep Medicine (AASM) as the repeated intermittent collapse of the upper (pharyngeal) airway causing interruption of airflow during sleep [5]. These interruptions cause oxygen desaturations and can lead to micro-arousals from sleep [5, 198]. OSAH can be largely asymptomatic [5]. When OSAH presents with symptoms, often Excessive Daytime Sleepiness (EDS), this is called Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS) [5, 198]. Throughout this thesis I shall refer to OSAHS.

OSAHS can be classified using a number of different measures. This thesis focuses on two of these. The first measure, the Apnoea-Hypopnoea Index (AHI), is a count of the number of apnoeas (complete cessations of breathing due to complete blockage of the airway) or hypopnoeas (periods of shallow breathing due to a partial collapse of the airway) lasting at least ten seconds that occur each hour [5, 198]. The AHI is an objective physiological measure of disease severity. Mild OSAHS is defined by the AASM as 5-15 events per hour⁸, moderate OSAHS as 15-30 events per hour, and severe OSAHS >30 events per hour [5].

The second measure of OSAHS is the extent of daytime sleepiness. Daytime sleepiness is generally quantified using the Epworth Sleepiness Scale (ESS) [113]. The ESS is a questionnaire completed by the patient rating the likelihood of falling asleep in eight different day-to-day situations in recent times on a scale from zero (would never doze) to three (high chance of dozing), giving a score between zero and 24 (Appendix A). An ESS score of less than 11 is considered normal for the general population [113, 114]. An alternative to the ESS is the Multiple Sleep Latency Test which measures physiologic sleep tendency by measuring how long it takes for an individual to fall asleep from the start of ‘nap-time’ [180]. Other measures of severity include Oxygen saturation nadir (the lowest value of oxygen saturation recorded in a sleep study) and the percentage of time asleep with oxygen saturation lower than 90% [122, 198]. These additional physiological values have not been considered, as

⁸An event is defined as an apnoea or hypopnoea

while the case study trial used in this thesis collected data on them they are not used in the CEA.

OSAHS is common in middle age, with risk factors including obesity, alcohol use and, potentially smoking [255, 257]. Men have twice the risk of developing OSAHS compared to women [198, 255]. OSAHS is causally linked with hypertension, leading to an increased risk of Cardiovascular Events (CVEs) including heart attacks and strokes [210]. Despite this link, due to alternative pathways which are not fully understood, the impact of treatment for OSAHS on Cardiovascular Disease (CVD) risk is still being explored [198]. Patients with OSAHS have an increased use of healthcare mainly due to their increased likelihood of CVEs [222]. Due to impaired vigilance, EDS leads to a two-to-three fold increase in risk of a Road Traffic Accident (RTA) [66]. Additionally, HRQoL is adversely affected by OSAHS through both increased rates of CVEs and the impact of EDS [134, 142].

OSAHS is thought to currently affect 2-7% of the adult population [173, 255]. Lee et al. (2008) summarised three population based studies, each finding the incidence of OSAHS was around 10% over a five year period [122]. However, Young et al. (1997) found over 80% of those with moderate-severe OSAHS and over 90% of those with mild OSAHS were undiagnosed [256]. The high non-diagnosis rate raises issues around the burden of OSAHS.

Current treatments for OSAHS

Two main treatments for OSAHS are currently available to patients in the UK [149]. Continuous Positive Airway Pressure (CPAP) involves the patient wearing a nasal or face mask connected to an electric air pump which generates pressure while asleep to try to prevent the upper airway collapsing [149, 198]. As an alternative to CPAP, particularly for those with mild OSAHS, Mandibular Advancement Devices (MADs) are often recommended, although not always available on the NHS [3, 149, 153, 198]. Many different types of MADs are available, each with the same basic idea - to hold the lower jaw and tongue forward, to try to maintain the upper airway during sleep [198]. Based on current evidence, MADs are less effective than CPAP but more effective than placebo [134, 198].

Current NICE recommendations state CPAP should be used for patients with moderate-severe OSAHS, with severity defined by the AHI [149]. For those with mild OSAHS, there is little evidence supporting treatment with CPAP [198]. However, NICE recommend CPAP for those with mild OSAHS who have exhausted all other treatment options [149]. MADs are an alternative to CPAP on the NHS in Scotland and a Cochrane review concluded MADs are an

appropriate alternative to CPAP for those who cannot tolerate or do not wish to use CPAP [3, 123]. The treatment of patients with mild OSAHS is addressed by the case study.

Other treatments for OSAHS are available. These include: lifestyle interventions (such as weight-loss, smoking cessation and reduction of alcohol intake); surgical options, which aim to increase the dimensions of the upper airway; and pharmaceutical treatments [133, 201, 219]. Additionally, NICE guidance on the insertion of implants into the roof of the mouth indicated while the procedure is safe there was no clinical evidence of its effectiveness [148]. Due to the lack of conclusive evidence on the effectiveness of these interventions they have not been considered in this thesis. However, lifestyle advice is the current best practice comparator in the case study for patients with mild-moderate OSAHS [198].

1.3.2 The Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea

The Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) was a RCT forming part of a NICE Technology Appraisal aiming to assess whether MADs are clinically beneficial and cost-effective, when compared to no treatment, in patients classified to have mild to moderate OSAHS on the basis of their AHI [175, 198]. Data and the results from TOMADO have been used extensively throughout the thesis.

TOMADO was an open-label, four-treatment, four-period, randomised, crossover trial which compared the clinical outcomes and cost-effectiveness of treatment with MADs and no treatment [175, 198]. Each treatment period lasted four weeks with a one-week washout period and a two-week acclimatisation period between treatments. The reduction of AHI was the primary outcome of the trial. EDS, measured by ESS, was an important secondary outcome. The trial considered three types of MADs differing in sophistication (bespoke, semi-bespoke and over-the-counter). A total of 90 patients were recruited to the trial, with 83 included in the final analysis. All patients had mild-moderate OSAHS (AHI: 5-30 events per hour, mean 13.8 events per hour (s.e. 6.2 events per hour)) and excessive daytime sleepiness (ESS ≥ 9 , mean 11.9 (s.e. 3.5)) [175, 198]. The population was 80% male, with an average age of 50.9 years (s.e. 11.6 years), and an average BMI of 30.6kgm^{-2} (IQR: $27.9 - 35.2\text{kgm}^{-2}$) [175, 198].

All three interventions were found to significantly decrease the AHI compared to no treatment (by 26%; 95% CI: 11-38% for the over the counter device; by 33%; 95% CI: 24-41% for

the semi-bespoke device; and by 36%; 95% CI: 24-45% for the bespoke device) [175, 198]. Similarly, all three interventions significantly reduced the ESS compared to no treatment (by 1.52; 95% CI: (0.73, 2.29) for the over the counter device; by 2.15; 95% CI:(1.31, 2.99) for the semi-bespoke device; and by 2.37; 95% CI:(1.53, 3.22) for the bespoke device) [175, 198]. The bespoke device was the most preferred treatment by the trial participants. The over the counter device had the greatest likelihood of discontinuation [175, 198].

The Health Technology Assessment (HTA) report concluded, for the TOMADO population, that the over the counter MADs appeared to improve patients' health and as the sophistication of the devices increased there were decreasing marginal returns in health improvement [198].

TOMADO was the first trial of the usage of a MAD for patients with mild to moderate OS-AHS containing clinical, patient centred, and cost-effectiveness outcomes [198]. It included a detailed study of HRQoL by collection of EuroQol 5 Dimension 3 Level scale (EQ-5D-3L) and Short Form 36 scale (SF-36) scores [27, 61, 237]. Disease specific Functional Outcomes of Sleep Questionnaire (FOSQ) and Short Calgary Sleep Apnoea Quality of Life Index (SAQLI) were also collected [75, 238]. These all found improvements in quality of life, compared to no treatment, with little difference between the interventions.

The data from TOMADO was used in updated meta-analyses models of the key outcomes (ESS, AHI and Systolic Blood Pressure (SBP)) as part of the HTA which helped to strengthen knowledge about the relative effectiveness of MAD and CPAP in patients with OSAHS [198–200].

A within-trial economic analysis, based on IPD from the four-week follow-up period in TOMADO, was carried out to assess the cost-effectiveness of the three interventions [175, 198]. The CEA compared each type of MAD in TOMADO and no treatment. All three MADs were found to be cost-effective, compared to no treatment, at a cost-effectiveness threshold, λ , of £20,000 per QALY gained [175, 198]. Very little difference in QALYs was found between the interventions so the incremental cost-effectiveness was driven by differing device costs. The semi-bespoke device compared to the over-the-counter device had an ICER of £186,844 per QALY gained and the bespoke device compared to the over-the-counter device had an ICER of £33,611 per QALY gained. The semi-bespoke device was dominated by the bespoke device. The semi-bespoke MAD was the most cost-effective intervention when $\lambda < £39,800$ per QALY gained, after which the bespoke MAD was the most cost-effective [198].

1.3.3 The lifetime CEA

OSAHS is a chronic condition associated with considerable long-term morbidities [134, 198]. Therefore, a lifetime CEA is required to gain a true insight into the cost-effectiveness of interventions for patients with OSAHS. A full description of the model can be found in Appendix B. This section provides an overview.

The McDaid et al. (2009) cost-effectiveness model

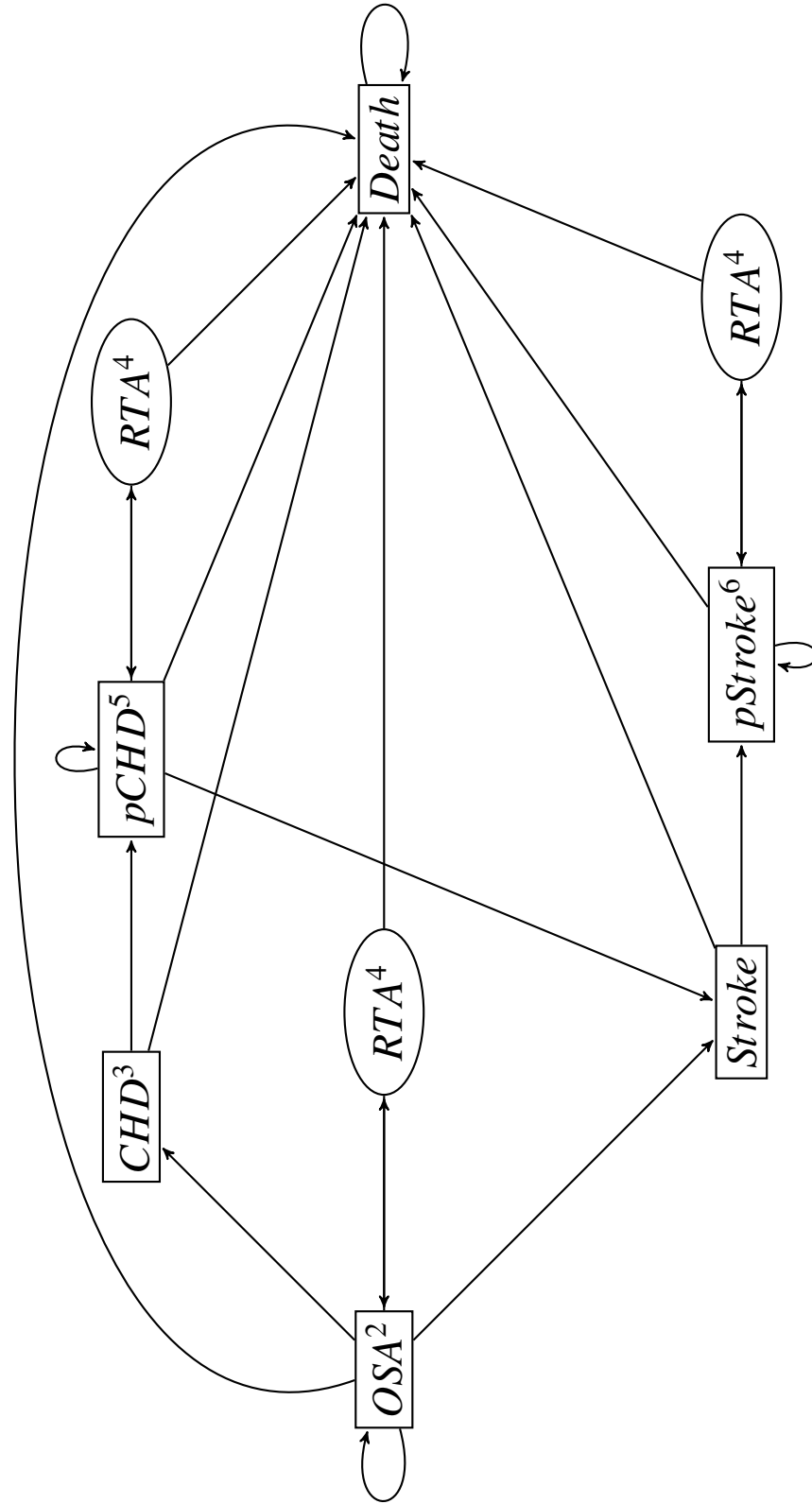
The lifetime CEA in Sharples et al. (2014), the case study, is based on work by McDaid et al. (2009) who developed a model (the McDaid model) to investigate the cost-effectiveness of CPAP, MADs, and Conservative Management (CM) (a one-off consultation with a GP, with some lifestyle advice on reducing or coping with symptoms) as part of a NICE Technology Appraisal [134, 149, 198]. Many different types of MADs and CPAP are available for use. However, due to small sample sizes, the CEA model groups all MADs and CPAP into one MAD or CPAP device group. [134]

The economic model is a state transition Markov model following the annual movement of a hypothetical homogeneous cohort of 50.5 year old men who were overweight (Body Mass Index (BMI) = 31.9 kg m^{-2}), had high blood pressure (SBP = 130mmHg), and an ESS of 11.9 [134]. This cohort was followed until almost all were dead, assumed to be a further 65 years⁹. The outcomes of the model were summarised as the incremental cost per QALY gained. The model included the possibility of a Coronary Heart Disease (CHD) event, stroke, and involvement in RTAs [134].

Figure 1.1 presents a simplified version of the model structure. All members of the cohort start in the Obstructive Sleep Apnoea (OSA) state and can remain there, unless an event occurs, until death. After a first non-fatal CHD event, individuals move into the post-CHD state (pCHD) reflecting increased morbidity and mortality due to a first non-fatal CHD event. Individuals can remain in the pCHD state until death, a RTA (fatal or non-fatal), or a stroke occurs. Second or subsequent CHD events are not modelled. After a non-fatal RTA, individuals return to their previous health state. After a first non-fatal stroke (possible from the OSA and pCHD states), individuals move to the post-stroke state (pStroke), where they remain until death or experience a RTA. Post-stroke CHD events are not modelled. Similar to the post-CHD state, the post-stroke state (pStroke) reflects the increased risk of mortality

⁹The time horizon is set so that the model is able to capture all potential differences in costs and outcomes [63].

Figure 1.1 A simplified schematic¹ of the lifetime cost-effectiveness model structure developed by McDaid et al. (2009) to explore interventions for patients with OSAHS [134]



¹ Patients are allowed to become non-adherent to their intervention during the first ten years. This simplified structure ignores the presence of adherent and non-adherent states and the transitions between them. ² All patients start in the OSA state; ³ CHD event state; ⁴ Non-fatal Road Traffic Accident; ⁵ Post-CHD event state; ⁶ Post-Stroke state

and morbidity due to a prior non-fatal stroke. As with CHD events, second or subsequent strokes are not modelled. A proportion of strokes are considered to be disabling, with these patients assumed unable to drive and so cannot have a RTA.

The McDaid model was populated using information from various sources, including systematic reviews, existing cost-effectiveness literature, the opinion of clinical experts and meta-analyses from the NICE Technology Appraisal [134, 149]. The transition probabilities in the model were derived from a number of sources. The costs included the cost of devices and on-going resource usage associated with maintenance and replacement of devices. Utilities and costs were assigned to each health state, some depending on the intervention. Uncertainty was explored using PSA (Section 1.1.2). The data in the model indicates a lack of robust evidence for some parameters. For example, for the risk of a RTA when using a MAD, RTA_{MAD} , the relationship:

$$RTA_{MAD} = RTA_{CPAP} \times \frac{\Delta ESS_{MAD-CM}}{\Delta ESS_{CPAP-CM}}$$

is used as opposed to direct data, where RTA_{CPAP} is the risk of an RTA using CPAP and ΔESS_{MAD-CM} and $\Delta ESS_{CPAP-CM}$ are the change in ESS due to treatment with MAD and CPAP respectively.

The McDaid model found CPAP was the most cost-effective intervention with a probability of 78% at a threshold of £20,000 per QALY gained [198]. MADs were more likely to be cost-effective for subgroups with more mild OSAHS [198].

The Sharples et al. (2014) cost-effectiveness model

The McDaid model was updated by Sharples et al. (2014) to reflect the TOMADO population and new evidence [134, 198]. The case study model had a similar population to the McDaid model, except the TOMADO population was slightly more overweight (BMI of 31.9 kgm^{-2} for TOMADO population compared to 30.0 kgm^{-2} in the McDaid model) [134, 198].

As in the case of the McDaid model, Sharples et al. (2014) used a variety of sources to populate the model [134, 198]. The parameters in the McDaid model were updated to reflect newly published data and the population having mild-moderate OSAHS [134, 198]. Costs were estimated in 2011/12 prices (£) [198]. Appendix B provides more information on the parameterisation of the CEA.

Changes to the cost-effectiveness model used in this thesis

This thesis has used the model developed by McDaid et al. (2009) and updated by Sharples et al. (2014) [134, 198]. The previous models were developed in Microsoft Excel. I have reproduced this model in R [9, 10, 45, 118, 139, 176, 245–248]. This allows for a more efficient computation of the model, through the use of parallelisation of the PSA simulations and ensures reproducibility [14]. In addition, value of information quantities (Section 1.1.3 and Chapter 2) apart from EVPI are essentially impossible to compute using Excel R [14].

There are a few changes to the CEA in this thesis compared to the Excel version used in Sharples et al. (2014) [198]. This is due to some small errors in the spreadsheet. These are:

- the Excel spreadsheet was not including those in the *pStroke* state adherent to their intervention in calculating life years
- for those treated with CM in the states corresponding to *pStroke* and *pCHD* and having had an RTA the cost of a RTA was not included.
- the utility decrement due to age was being incorrectly applied to those who had an RTA, CHD event or a Stroke event.
- a smaller number of PSA samples have been used in the Excel spreadsheet, meaning results are presented to a lesser degree of accuracy.

1.3.4 Results from the lifetime cost-effectiveness analysis

The case study lifetime CEA, fully illustrated in Appendix B, estimated that MADs were more expensive and more effective than CM (Table 1.2) [198]. Additionally, CPAP was estimated to be more expensive and more effective than MADs. These results are presented using 500,000 PSA simulations.

Both Figure 1.2 and Table 1.2 show at traditional thresholds (around £20,000 per QALY gained) CPAP is most likely to be cost-effective. However, there is a significant amount of uncertainty around this decision suggesting potential value of further research. The CEAF (Figure 1.3) shows uncertainty around the optimal treatment decision at traditional thresholds.

Table 1.2 Results from the case study cost-effectiveness analysis using 500,000 PSA samples calculated in R

	CM	MAD	CPAP
Total Expected LYs ¹	28.36	28.51	28.61
Total Expected QALYs	14.35	14.65	14.68
Total Expected cost (£)	6,112	8,331	8,501
NMB (£) ²	280,888	284,669	285,099
ICER (£ per QALY gained)		7,397 ³	5,667 ⁴
Pr(Cost-effective at threshold values)			
£10,000 per QALY gained	0.19	0.40	0.41
£20,000 per QALY gained	<0.01	0.44	0.56
£30,000 per QALY gained	<0.01	0.41	0.59

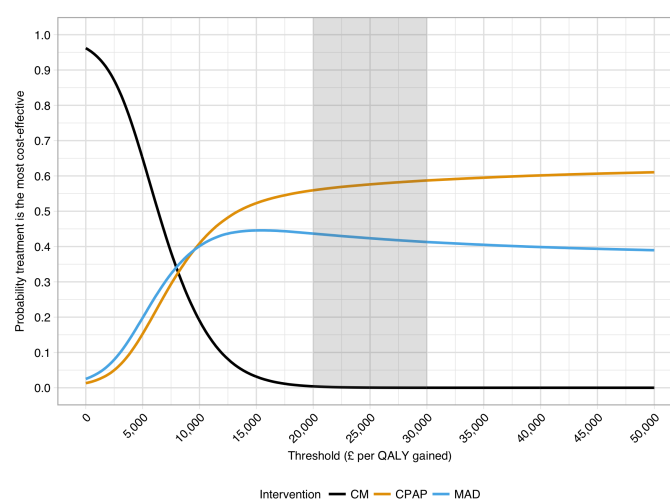
¹: Uncertainty is presented as the probability of the intervention being cost-effective rather than an interval as per Claxton (1999) [37]. In this paper it states confidence intervals are not appropriate in CEA as the objective is to maximise health gains. Therefore, decisions should be based on expected cost-effectiveness given existing information. One of the alternatives must be chosen. *"The opportunity cost of failing to make the correct decision based on expectation is symmetrical and the historical accident that dictates which of the alternatives is regarded as 'current practice' is irrelevant"* [24, 37].

²: NMB calculated using a cost-effectiveness threshold λ of £20,000 per QALY gained.

³Compared to CM

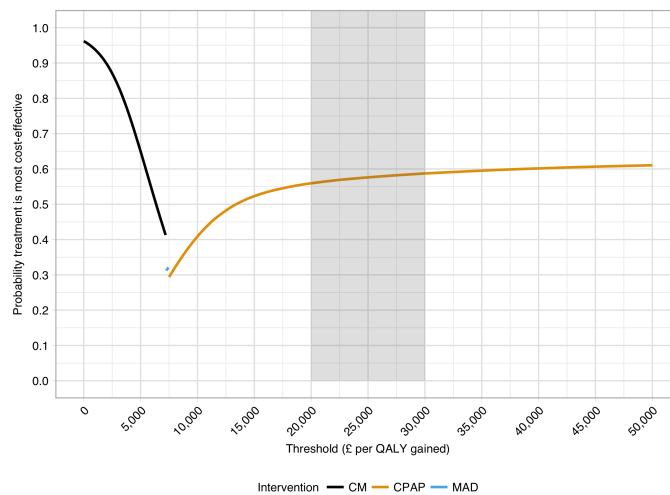
⁴Compared to MAD

Figure 1.2 The cost-effectiveness acceptability curve for patients with OSAHS treated with CM, MAD and CPAP for a range of thresholds for the case study cost-effectiveness analysis using 500,000 PSA samples¹



¹ The shaded area represents the usual cost-effectiveness thresholds region, as used by NICE

Figure 1.3 The cost-effectiveness acceptability frontier for patients with OSAHS treated with CM, MAD and CPAP for a range of thresholds for the case study cost-effectiveness analysis using 500,000 PSA samples¹



¹ The shaded area represents the usual cost-effectiveness thresholds region, as used by NICE

1.3.5 Recommendations for future work

The case study report presents a list of future research priorities for OSAHS [198]. These include: the collection of further information and data on long-term compliance; studies on HRQoL over time; clinical and cost-effectiveness comparisons for the different types of MADs over a range of severities; head-to-head comparisons of MAD and CPAP in mild OSAHS; and further data on long-term risk of CVD and its risk factors.

1.4 Thesis aims and objectives

The overall aim of this thesis is to explore methods of quantifying uncertainty and targeting where reducing uncertainty would be most beneficial in CEAs. This involves using modelling techniques to help inform better treatment decisions. All theoretical work is applied to the case study CEA (Chapter 1.3.3 and Appendix B). This thesis focuses on two under explored aspects of CEAs: patients' adherence to interventions and heterogeneity between patients.

Objective 1: Exploring the impact of modelling adherence to interventions on the results of the CEA and the value of collecting more information on adherence to interventions

The first objective of the thesis is to develop methods for modelling patients' adherence to an intervention and the importance of this for decision models. It is important that adherence to an intervention is reflected in cost-effectiveness models to replicate real-life use.

Within this objective there are two main research targets. The first is to develop innovative methods to model adherence to interventions using all available information which can be used as part of the health economic model. The results from these models can be used to assess the impact of adherence on the optimal treatment decision and its uncertainty. These methods are applied to the case study CEA. The second research target is to extend methods to assess the value of collecting further information on adherence to interventions. Special attention is paid to the choice of timepoints at which data on adherence should be collected.

Objective 2: Exploring the stratification of the optimal treatment decision based on non-binary measures of disease severity and the value of collecting more information to guide stratification

If CEAs are used to make decisions based on whether an intervention is cost-effective for the population on average there may be groups of the population who receive a suboptimal intervention. This section of the thesis uses evidence synthesis methods to assess whether treatment should be stratified based on single or multiple measures of disease severity. A second research target is to develop and apply methods for identifying the populations where the collection of further information on the treatment effect would be most beneficial to guide decisions on stratification.

1.5 Structure of the thesis

The thesis is structured as follows:

Chapter 2 begins by revisiting the value of information quantities outlined in Section 1.1.3. A range of current methodologies for the calculation of these quantities are presented with their advantages and disadvantages. The importance of the PSA sample size for the accuracy of the value of information estimates is illustrated using some of the current methodologies

applied to the case study. Additionally, methods for the calculation of population-level value of information quantities and the associated challenges are described and applied.

Chapter 3 explores adherence to interventions. It begins by summarising current guidance in incorporating adherence to interventions in CEAs. Results from a literature review extracting data on long-term adherence to treatment for patients with OSAHS are reported. It then develops a Bayesian meta-analysis model for synthesising the adherence data found in the literature review. The impact of the results of these methods on the case study CEA is explored along with an analysis of the expected value of perfect partial information (EVPPI) and the expected value of obtaining specific further data (EVSI) on adherence.

Chapter 4 explores current and new Bayesian meta-regression methodology for combining aggregate and individual participant data to stratify the optimal treatment decision when considering non-binary measures of disease severity. Studies which provided information in the case study HTA on the treatment effect were reviewed and more detailed information on baseline characteristics and any individual participant data were extracted. This data is used in a Bayesian network meta-regression model exploring potential stratification on one or two covariates of interest. The impact of the network meta-regressions on the CEA are explored to identify any benefit of stratification on an individual patient-level.

Chapter 5 examines the population-level value of stratification to identify whether a stratified treatment regime should be implemented using the results of Chapter 4. Focus is given to the distribution of the stratifiers in the population, the adherence to the stratified regime, and the costs involved with the implementation of a stratified regime. The value of information methods introduced in Chapter 2 are used and extended to heterogeneous populations and to prioritise further research to guide stratification.

Chapter 6 concludes the thesis, focussing on its contributions to quantifying uncertainty and targeting future research. This chapter ends with a discussion of future research questions that emerge from this work.

Chapter 2

Estimating value of information quantities

Value of information quantities, introduced in Section 1.1.3, describe the expected value of collecting further information to reduce uncertainty in a CEA. Historically, these quantities have had a large calculation burden which has impacted on their application. In recent years a number of more efficient methods for estimating the expected value of perfect partial and sample information have been published. This chapter summarises some of these methods for use later in the thesis. Even less focus is given in the literature to the errors associated with the estimation of value of information quantities. Errors are present when estimating value of information quantities using all estimation methods, due to a finite PSA sample size. This Chapter presents a method for calculating these errors for EVPI and EVPPI when estimated using non-parametric regression. Uncertainty around quantities such as disease prevalence and incidence, intervention take-up rate, and an appropriate time horizon for technologies needed to scale per person value of information quantities to a population-level is also frequently neglected. The impact of changes to these values is investigated through an application to the case study CEA.

2.1 Introduction

A CEA provides decision makers with information as to whether an intervention is cost-effective on current evidence. However, it does not routinely include information on whether collecting further information is likely to alter the treatment decision.

As outlined in Section 1.1.3 when deciding whether an intervention is cost-effective, two main questions need to be answered [24, 165]:

- Under current information, should the new intervention be recommended?
- Should further information be collected to reduce uncertainty around the optimal treatment decision?

Under a Bayesian decision-theoretic framework, an intervention is accepted as cost-effective based on its expectation but the potential value of collecting further information is still important [37]. If a wrong decision is made, costs are incurred. These may be monetary costs. For example, in terms of extra costs due to treatment or adverse events [24, 37]. However, they can also include the costs of the health foregone to the population by making a suboptimal decision [24]. Analysis using current information contains uncertainty through the parameters of the CEA, θ , which are often estimated from a variety of sources and rarely known precisely. Additionally, there may be uncertainty around other aspects of the CEA model, such as its structure. The focus of this chapter is on parametric uncertainty. Section 2.7 briefly discusses other forms of uncertainty in a CEA model.

Value of information quantities provide a way to estimate the expected health gains from collecting further information. They use both the opportunity cost of making a wrong decision and the likelihood of making a wrong decision [24]. In this way, decision makers can see the ‘penalty’ of making a wrong decision [24]. A low penalty, because of a low opportunity cost and/or a low probability of making a wrong decision indicates little benefit in further analysis and so the optimal treatment decision could be made using current evidence. The combination of the opportunity cost and the likelihood of a wrong decision is important [12, 24]. Even if the probability of making a wrong treatment decision is high, there is only value in collecting further information if there is a cost to this decision [12, 24].

The results from value of information analyses can be used to set future research priorities. This can ensure further research is focussed on areas where uncertainty in the model is of greatest consequence to the decision and hence the greatest value to the population. This is particularly important when resources for future research are scarce.

Currently, few applied studies implement value of information analysis. Steuten et al. (2013) carried out a systematic review of the methods and applications of value of information in published literature [207]. They found that, although new methods of calculation were being published, the number of applications remained low. Focus groups¹ carried out by Bindels et al. (2016) explored why value of information calculations were not carried out [20]. They

¹Consisting of researchers, policy makers, and representatives of pharmaceutical companies.

found reasons included the inability of value of information quantities to easily incorporate all uncertainties within a model (for example, structural uncertainty) and the complex, often time consuming, calculations required [20].

This chapter highlights the importance of calculating value of information quantities by building on the concepts introduced in Section 1.1.3. The first aim of this chapter is to provide information on three different value of information quantities - EVPI, EVPPI, and EVSI and some of the recently published methods which enable more efficient calculations of these quantities. Some of these methods are applied to the case study CEA to motivate areas of interest for further research.

A second aim is to present and illustrate methods to calculate the error around the value of information estimates and its relationship to the number of PSA simulations. It is important to consider the accuracy of the estimates which can impact on calculation time, the CEA results, and whether future research should be implemented. Methods for calculating the error around the EVPPI and EVSI estimates obtained from a commonly used non-parametric regression procedure are presented. An extension to include the error from a finite PSA sample size in addition to error from the non-parametric regression is presented and illustrated.

A third aim is to present population-level values for the value of information quantities. The population-level value of collecting further information is compared to the costs of further research to assess whether research should be carried out in practice. A sensitivity analysis to the case study CEA explores the impact of uncertainty around quantities such as the incidence and prevalence of OSAHS used to estimate the population-level EVPI.

The chapter will proceed as follows. Firstly, Section 2.2 outlines the theory behind EVPI and how it can be calculated at an individual and population-level. The theory behind the calculation of standard error associated with estimating the EVPI from a limited number of PSA samples is presented. Section 2.3 outlines the theoretical framework for the calculation of EVPPI along with an overview of a number of new methods developed to assist its calculation. Section 2.4 presents an overview of the theory behind EVSI and outlines recently developed methods for its calculation. The method for calculating the standard error for the EVPPI estimator based on non-parametric regression is extended in Section 2.5 in order to quantify Monte Carlo error from a finite number of PSA samples in addition to error from the regression coefficients. Section 2.6 applies methods introduced earlier in the chapter to the case study CEA to find an optimal number of PSA samples to achieve value of information

quantities to a sufficient level of accuracy. The EVPPI values and their standard errors are presented for a number of parameters in the case study CEA. The population-level EVPI is presented including a sensitivity analysis on the impact of (modest) changes to treatment uptake, diagnosis rate, population size, and time horizon. Section 2.7 provides a discussion of the calculation methods and why value of information is rarely calculated along with future research priorities. Section 2.8 concludes the chapter with details on how this work is used throughout the thesis.

2.2 The Expected Value of Perfect Information

The EVPI is the upper limit for the amount of money a decision maker should be willing to pay for research that would eliminate uncertainty in all parameters, θ , in a CEA [24].

2.2.1 Background and theory

The basic premise behind EVPI was outlined in Section 1.1.3. To re-cap, if perfect information existed for θ the correct decision on the optimal treatment j would always be made [24]. Three main components determine the magnitude of the EVPI [24, 37]:

1. how cost-effective the current optimal treatment is
2. the level of uncertainty around the cost-effectiveness decision
3. the consequences of a wrong decision, i.e. the penalty of making a wrong decision

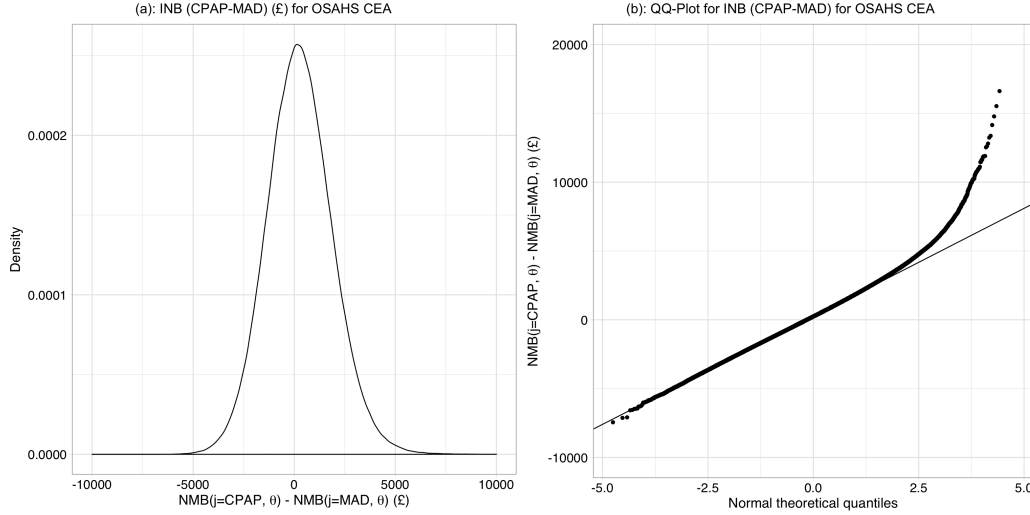
Analytical calculation of EVPI

If the INB can be assumed to be normally distributed, the EVPI can be calculated analytically. This method is used in earlier value of information analyses and presented in Wilson (2014) [37, 38, 40, 251]. It is more frequently used in trial-based analyses as opposed to model-based CEAs [251]. When data for the CEA model come from a number of sources the INB is less likely to have a normal distribution, or may not fit any parametric distribution. Therefore, simulation based approaches are preferable [24]. Figure 2.1 shows the distribution of the INB between CPAP and MAD for the (model-based) case study CEA and the associated QQ-plot. The QQ-plot in particular highlights the non-normality of the INB in this example.

Simulation approach to calculating the EVPI

While the EVPI can sometimes be calculated analytically, it is simple to estimate the EVPI using the PSA output [24]. Recalling Section 1.1.3, the EVPI is the difference between the

Figure 2.1 The Incremental Net Benefit for the MAD vs CPAP treatment comparison and QQ-plot for normality 100,000 PSA samples and a cost-effectiveness threshold of £20,000 per QALY gained



expected NMB under perfect versus current information. Let $NB(j, \theta)$ represent the NMB for intervention j with parameter values θ . The EVPI is:

$$EVPI = E_{\theta} \left[\max_j NB(j, \theta) \right] - \max_j E_{\theta} [NB(j, \theta)] \quad (2.1)$$

This can be estimated from K PSA samples using the Monte Carlo mean for each expectation with respect to θ as:

$$\widehat{EVPI} = \frac{1}{K} \sum_{k=1}^K \left[\max_j NB(j, \theta^{(k)}) \right] - \max_j \left[\frac{1}{K} \sum_{k=1}^K NB(j, \theta^{(k)}) \right] \quad (2.2)$$

The equivalence of the EVPI and the expected opportunity loss

The EVPI is equivalent to the expectation of the opportunity loss [24]. The opportunity loss, $L(\theta)$, associated with making an incorrect decision is:

$$L(\theta) = \max_j NB(j, \theta) - \max_j E_{\theta} [NB(j, \theta)]$$

This is the difference between the NMB for the optimal intervention and the alternative under current information. The expected opportunity loss, $E_{\theta} [L(\theta)]$ can be expressed as:

$$E_{\theta} [L(\theta)] = E_{\theta} \left[\max_j NB(j, \theta) - \max_j E_{\theta} [NB(j, \theta)] \right] \quad (2.3)$$

which is equivalent to Equation 2.1.

2.2.2 Population-level Expected Value of Perfect Information

Equation 2.1 is the EVPI around making the decision for a particular individual or for a particular patient episode [24]. Any future research producing additional evidence can benefit all current and future patients². Therefore, it is important to consider the EVPI for the population who could benefit from the additional evidence both now and in the future [24].

As introduced in Section 1.1.3, calculation of population-level EVPI requires additional pieces of information: the expected future lifetime of the technology (T), an estimate of the rate of incidence of the disease over T (I_t), the current prevalence rate of the disease (I_0), the size of the decision population at risk (P), and a discount rate (i) - typically 3.5% per annum in England and Wales (outlined in the NICE Methods Guidance [150]). Let $t = 1, \dots, T$ index years assuming data is available on an annual basis. The population-level EVPI (popEVPI) is:

$$popEVPI = EVPI \times P \times \left(I_0 + \sum_{t=1}^T \frac{I_t}{(1+i)^t} \right) \quad (2.4)$$

If the cost of research is less than the population EVPI, it is a necessary but not sufficient condition for the implementation of further research.

There are difficulties in estimating the values required to calculate the population EVPI. Estimating the appropriate time horizon for the population EVPI is challenging. Philips et al. (2008) found finite time horizons for decision problems are a proxy for future technology change [165]. The impact of changing these quantities on the population-level EVPI is explored in Section 2.6.3.

The uptake of an intervention can impact on the population-level EVPI [86]. The uptake of the most cost-effective intervention under current information is a measure of the proportion of the eligible population who actually use the intervention [86]. If this is less than 100%, the population-level EVPI can be over-estimated. This can lead to suboptimal spending of resources and delays to patients receiving their optimal treatment. Grimm et al. (2015) found many studies implicitly ignore uptake of interventions in calculating their population-level EVPI. Additionally, very few studies highlighted that population-level EVPI was subject to

²For non-chronic diseases, those patients who were part of the additional research will not benefit from the research in future.

uncertainty in the population estimates (P, I_0, I_t, T) [86].

This idea is akin to the value of implementation which deals with the assumption that CEAs assume the most cost-effective intervention is implemented [73, 249]. Less than perfect implementation of cost-effectiveness guidance leads to suboptimal resource allocation incurring costs and impacting on the health of the population [73, 103].

There may be evidence the uptake rate may change over time. Uptake rate can be incorporated into the population-level EVPI. Let U_t be the uptake rate of the intervention at time t then:

$$popEVPI = EVPI \times P \times \left(I_0 \times U_0 + \sum_{t=1, \dots, T} \frac{I_t U_t}{(1+i)^t} \right) \quad (2.5)$$

This assumes the uptake rate for all interventions is the same. Uptake rates that differ between interventions cannot be modelled in this way. Grimm et al. (2015) found no study that highlighted the uptake rate may be dynamic in nature [86].

In a similar way, Equation 2.5 can model diagnosis rates by replacing U_t with D_t - the proportion of the population with the disease who are diagnosed. Only those diagnosed can benefit from further research and/or use the intervention. It is feasible the diagnosis rate may change over time. For example, due to improved diagnostic tests or increased public awareness.

In addition, Ades et al. (2004) state that the size of the population who enter the population-level EVPI model each year can change over time [2]. This population size can be defined for each $t = 1, \dots, T$ as $Q_t = P_t \times I_t \times U_t \times D_t$. Therefore, Q_t can incorporate uptake rates, diagnosis rates, changing population size, and disease incidence rates.

There is a similar term leakage which is discussed further in Chapter 5 [47]. Leakage is the proportion of participants receiving the most cost-effective intervention as opposed to an alternative intervention [47]. This is subtly different to the idea of uptake. Under imperfect uptake those who do not use the intervention do not receive an alternative, whereas under leakage, patients could receive a suboptimal intervention. Leakage impacts on the value of stratification.

2.2.3 Monte Carlo error

PSA involves repeated sampling of the distributions of the uncertain parameters, $\boldsymbol{\theta}$ [12, 13, 24]. An insufficient number of PSA samples can lead to incorrect decisions being made should the results not have converged to an adequate degree of accuracy. The error due to the number of simulations is called the Monte Carlo error.

Suppose there are K posterior samples, $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)}$, for a generic function $g(\boldsymbol{\theta})$. The Monte Carlo estimate of $E_{\boldsymbol{\theta}}[g(\boldsymbol{\theta})]$ using the K samples is:

$$\bar{g}_K = \frac{1}{K} \sum_{k=1}^K g(\boldsymbol{\theta}^{(k)}) \quad (2.6)$$

If $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)}$ are independent, the Central Limit Theorem can be applied to give the Monte Carlo Standard Error (MCSE) for \bar{g}_K :

$$\begin{aligned} MCSE &= \sqrt{\text{Var}(\bar{g}_K)} \\ &= \sqrt{\frac{\text{Var}(g(\boldsymbol{\theta}))}{K}} \end{aligned} \quad (2.7)$$

where the posterior variance $\text{Var}(g(\boldsymbol{\theta}))$ can be estimated using the empirical variance of the K samples thus the MCSE can be estimated by:

$$\widehat{MCSE} = \frac{s}{\sqrt{K}} \quad (2.8)$$

with

$$s^2 = \frac{1}{K-1} \sum_{k=1}^K \left[g(\boldsymbol{\theta}^{(k)}) - \bar{g}_K \right]^2 \quad (2.9)$$

While this error is important in the context of CEAs, current guidance recommends CEA models should be run until convergence, with no definition of convergence provided [150]. Some work has been carried out to investigate the optimal number of simulations required for a PSA looking at the convergence of costs, QALYs, ICERs, and NMBs [91]. This found the current convention in HTAs of 1,000 PSA simulations was insufficient, and suggested a broad estimate of 10,000 PSA simulations was required to estimate these values to a reasonable degree of accuracy [91]. However, the optimal number of simulations will in practice be

highly dependent on model in question.

In a CEA and in calculating the EVPI a draw from the distribution for each uncertain parameter is taken for each simulation, meaning Monte Carlo error is present in the EVPI estimate. It is important to be aware of this error to ensure correct decisions are made on further research. As EVPI is an upper bound to the potential value of further research, it is rarely necessary to know the EVPI to the nearest £1. Learning the EVPI to a lesser degree of accuracy can be useful, for example to compare the uncertainty across different CEAs.

To calculate the Monte Carlo error, the EVPI needs to be reformulated in terms of the expected opportunity loss from making the wrong decision (Equation 2.3). The MCSE for the EVPI estimate from K PSA samples can be calculated using Equations 2.6 - 2.9 setting:

$$g(\boldsymbol{\theta}^{(k)}) = L(\boldsymbol{\theta}^{(k)})$$

and the MCSE can be estimated as $\frac{s}{\sqrt{K}}$ where s is the empirical variance of $L(\boldsymbol{\theta}^{(k)})$.

The approximate width of the 95% Monte Carlo interval is $2 \times 1.96 \times MCSE$. This width can be used to guide what K should be by going backwards from the desired degree of accuracy. Alternatively, it can indicate how many significant figures are appropriate to present estimates using K PSA samples. This is particularly useful when K is limited due to computational complexity, for example. Oakley et al. (2010) outline a similar approach for calculating the width of the 95% confidence interval for a two-level Monte Carlo estimate for EVPPI [154].

2.3 The Expected Value of Perfect Partial Information

2.3.1 Background and theory

The EVPI (Sections 1.1.3 and 2.2) provides an upper bound for the value of research which would eliminate uncertainty in all parameters, $\boldsymbol{\theta}$, in the CEA. It is unlikely uncertainty in all parameters can be eliminated, so the EVPI is generally regarded a ‘theoretical quantity’. The expected value of perfect partial information or the expected value of perfect information for parameters (EVPPI) is the value of eliminating uncertainty for a subset of parameters, $\boldsymbol{\phi} \subseteq \boldsymbol{\theta}$. The EVPPI can help focus future research priorities by identifying parameter(s) where more precise estimates will be of most value. The EVPPI is also regarded as a ‘theoretical quantity’.

An infinite sample size would be required to eliminate uncertainty in any set of parameters.

As outlined in Section 1.1.3, the EVPPI is the difference between the expected value with perfect and current information on the parameter(s) of interest [12, 23, 24, 154]. Formally, the expected value of perfect information for a subset of parameters (ϕ), where θ is the set of all uncertain parameters, ($\phi \subseteq \theta$, $\phi \cup \bar{\phi} = \theta$) for a set of j interventions is [12, 23, 24, 154]:

$$EVPPI_{\phi} = E_{\phi} \left[\max_j E_{\bar{\phi}|\phi} [NB(j, \phi, \bar{\phi})] \right] - \max_j E_{\theta} [NB(j, \theta)] \quad (2.10)$$

2.3.2 Methods of calculation

Monte Carlo methods

The traditional method for calculating EVPPI uses two-level Monte Carlo samples with the second term in Equation 2.10 estimated using a single-level Monte Carlo simulation [23, 154]: In this method $\theta^{(1)}, \dots, \theta^{(K)}$ are sampled from the distribution of θ and $NB(j, \theta^{(k)})$ calculated for all j and $k = 1, \dots, K$, i.e. the output from a PSA is used to calculate:

$$NB^* = \max_j E_{\theta} [NB(j, \theta)]$$

and by using a Monte Carlo estimator over θ is estimated as:

$$\widehat{NB}^* \approx \max_j \frac{1}{K} \sum_{k=1}^K NB(j, \theta^{(k)}) \quad (2.11)$$

For a sufficiently large K , Equation 2.11 is equivalent to calculating the optimal intervention in terms of NMB from a PSA (Section 1.1.2). The first term in Equation 2.10 causes computational difficulties due to the nested expectations. Using notation from Oakley et al. (2010) [154], let:

$$m(\phi) = \max_j E_{\bar{\phi}|\phi} [NB(j, \phi, \bar{\phi})]$$

The first term in Equation 2.10 is $E_{\phi} [m(\phi)]$. In general, $m(\phi)$ cannot be calculated analytically. Monte Carlo sampling methods can be used to estimate $m(\phi)$ [154]. First, a value of ϕ is sampled from its joint distribution. Then, L values of $\bar{\phi}$ are sampled from $\bar{\phi}|\phi$, $\{\bar{\phi}^{(1)}, \dots, \bar{\phi}^{(L)}\}$. The CEA model is run for each of the L samples giving $NB(j, \phi, \bar{\phi}^{(l)})$ for all j and $l = 1, \dots, L$. $m(\phi)$ can be estimated by:

$$\hat{m}(\phi) \approx \max_j \frac{1}{L} \sum_{l=1}^L NB(j, \phi, \bar{\phi}^{(l)})$$

$E_{\phi}[m(\phi)]$ can be approximated by $E_{\phi}[\hat{m}(\phi)]$. This is repeated N times with $\phi = \{\phi^{(1)}, \dots, \phi^{(N)}\}$ simulated from their distributions. For each $n = 1, \dots, N$, $\hat{m}(\phi^{(n)})$ can be calculated and using a Monte Carlo estimate:

$$\hat{E}_{\phi}[\hat{m}(\phi)] = \frac{1}{N} \sum_{n=1}^N \hat{m}(\phi^{(n)})$$

Therefore, the EVPPI can be estimated as:

$$\begin{aligned} \widehat{EVPPI}_{\phi} &= \hat{E}_{\phi}[\hat{m}(\phi)] - \widehat{NB}^* \\ &= \frac{1}{N} \sum_{n=1}^N \max_j \left[\frac{1}{L} \sum_{l=1}^L NB(j, \phi^{(n)}, \bar{\phi}^{(n,l)}) \right] - \max_j \left[\frac{1}{K} \sum_{k=1}^K NB(j, \theta^{(k)}) \right] \end{aligned} \quad (2.12)$$

where $N \times L = K$ and $\bar{\phi}^{(n,l)}$ is the l^{th} sample drawn from $\bar{\phi} | \phi = \phi^{(n)}$. As in Section 2.2.3, Monte Carlo estimates contain error due to random sampling, with more samples reducing the uncertainty, so L and N need to be sufficiently large [154]. $\hat{E}_{\phi}[\hat{m}(\phi)] - \widehat{NB}^*$ is a biased estimator of $EVPPI(\phi)$ [154]. This bias is independent of N , but depends on L , meaning both L and N need to be sufficiently large to reduce uncertainty and bias to an appropriate level. Oakley et al. (2010) found for their case studies $L = N = 500$ gave a sufficiently reasonable estimate for EVPPI. However, these models were considerably simpler than the case study in this thesis [154]. They also highlight decreasing marginal returns due to increasing L and N [154]. As an example, increasing L and N from 500 to 1,000 increases the number of simulations and computational time four fold, with the resulting increase in accuracy of the estimate being much smaller [154]. Oakley et al. (2010) ignored Monte Carlo error due to finite $K (= L \times N)$ in assessing the optimal K [154]. Section 2.5 addresses this error in the context of other methods.

The computational burden of estimating EVPPI using two-level Monte Carlo estimation is one of the reasons why EVPPI is not normally carried out in practice. Further, standard health economics textbooks, such as Briggs et al. (2006), only present this method along with simplifications if the NMB has a linear relationship between ϕ and the $NB(j, \theta)$ [24, 241]. A reduction in the computational burden for EVPPI is especially pertinent when researchers

want to consider different sets of ϕ to prioritise future research.

A number of methods to estimate EVPPI have recently been developed focussing on accuracy and reducing computational expense. The remainder of this section reviews and appraises methods for calculating and estimating EVPPI. These include both general purpose methods and methods only applicable in special cases. Some of these methods and a comparison in terms of their estimation abilities and their standard errors are outlined in a paper by Heath et al. (2017) [95, 187, 212, 215]. Heath et al. (2017) found little difference in the speed and accuracy of the Sadastafavi (2012), Strong and Oakley (2013), and Strong et al. (2014) calculation methods [95, 187, 214, 215]. Additionally, the Heath et al. (2016) EVPPI method was found to be more efficient than the Strong et al. (2014) method, as presented in Heath et al. (2016) [93, 215]. All of these methods showed marked improvements in calculation times compared to the two-level Monte Carlo methods. Since the two-level Monte Carlo method was feasible in the examples in the papers presenting the new methods, the EVPPI could be determined accurately to arbitrary precision. Thus, all new methods compared their results to these values in terms of accuracy and found comparable agreement.

In addition to the methods reviewed in Heath et al. (2017), two further methods are outlined in the remainder of this section - Jalal et al. (2016) and Heath et al. (2016) [93, 95, 108]. A summary of the situations where each method can be used and their relative merits are presented in Table 2.1.

Calculation of the Expected Value of Perfect Partial Information in special situations

There are some situations where the EVPPI can be calculated analytically, for example: if the INB can be assumed to have a normal distribution, when the NMB can be re-parametrised to be linear, and when a Taylor approximation can be used for the net benefit function [24, 128, 251].

The incremental net benefit can be assumed to have a normal distribution

This method is generally applied to trial based economic evaluations, although can also be applied to model based economic evaluations when the normality assumption holds [251]. Wilson (2014) outlined how to calculate EVPPI when the NMB is normally distributed, or it is reasonable to assume normality, using the same approach as the analytic calculation of EVPI (Section 2.2.1) [251].

Table 2.1 Reviewed studies for the methods estimating the Expected Value of Perfect Partial Information along with their advantages and disadvantages

Method	dim ϕ^1	Summary	Advantages	Disadvantages
Two-level Monte Carlo [23, 154]	1+	Two-level Monte Carlo sampling	Gives answer correct to the required precision as the sample size increases	Computationally intensive; biased estimator; conditional distribution of $\phi \bar{\phi}$ only has a closed form if ϕ and $\bar{\phi}$ are correlated
Analytic Calculation [251]	1+	Analytic calculation	Quicker than Monte Carlo based methods; can be calculated in Excel	Assumes INB is normally distributed
Madan et al. (2014) [128]	1+	Approximation methods for the conditional expectation in EVPPI when the INB has a specific form using one-level Monte Carlo	Quicker than two-level Monte Carlo sampling methods	Only works for specific functional forms of $NB(j, \theta)$
Strong and Oakley (2012) [212]	1	Ordered one-level Monte Carlo algorithm	Quicker than two-level Monte Carlo sampling methods; applicable to any scenario where no analytic solution for calculating EVPPI exists	Can only be used for one parameter of interest; size of subsets within the model can be difficult to estimate
Sadatsafavi et al. (2013) [187]	1	Ordered one-level Monte Carlo algorithm	Quicker than two-level Monte Carlo sampling methods; can be used for individual-level microsimulation models	can only be used for one parameter of interest; can be sensitive to how the samples are partitioned
Strong et al. (2014) [215]	1+	One-level Monte Carlo using non-parametric regression using Gaussian Process (GP) or Generalised Additive Models (GAM)	Quicker than two-level Monte Carlo sampling methods; readily available applications for calculation; can be used for correlated parameters of interest; Can be used for patient-level models	GAM specification increasingly difficult to fit as the number of parameters increases; GP specification requires user interaction to ensure convergence occurs

Method	$\dim \phi^1$	Summary	Advantages	Disadvantages
Heath et al. (2016) [93]	2+	Non-parametric regression using Integrated Nested Laplace Approximation and Principal Fitted Components	Quicker than two-level Monte Carlo sampling methods; readily available applications for calculation	Only works for two or more parameters of interest; small loss in accuracy compared to Strong et al. (2014) when ϕ is large [215]
Jalal et al. (2015) [108]	1+	Unit Normal loss integrals and linear regression meta-modelling	Quicker than two-level Monte Carlo sampling methods	Restricted by the prior distribution of parameter(s) of interest needing to be approximately normal; restricted by the normality assumption for the relationship between the INB and parameter(s) of interest; difficult to calculate when three or more interventions

¹The number of parameters in ϕ required for the method to be viable.

This analytic method is computationally quicker than two-level Monte Carlo methods and is not subject to Monte Carlo error, a feature of simulation based methods. It can be calculated in Microsoft Excel, with the author providing Excel workbooks [251]. However, it relies on the assumption the INB is normally distributed. Normality is unlikely in situations where (complex) CEA models are used, as illustrated by the OSAHS case study (Figure 2.1). However, if a trial based economic evaluation is used the Central Limit Theorem can be applied to the means making the assumption the INB has a normal distribution more reasonable.

Other special situations

Madan et al. (2014) showcase a range of methods for estimating the EVPPI in specific situations [128]. They use the EVPPI re-expressed as:

$$EVPPI(\phi) = E_{\phi} \left[\max_j \left\{ E_{\bar{\phi}|\phi} [NB(j, \theta) - NB(j^*, \theta)] \right\} \right] \quad (2.13)$$

where j^* is the intervention maximising NMB. All methods in Madan et al. (2014) involve expressing $E_{\bar{\phi}|\phi} [NB(j, \theta) - NB(j^*, \theta)]$ in an analytic form meaning $EVPPI(\phi)$ can be estimated using a single Monte Carlo simulation [128].

The first set of methods can be used when the expectations for $\bar{\phi}$ can be plugged into $EVPPI(\phi)$ (Equation 2.13). If $NB(j, \theta)$ can be expressed as:

$$NB(j, \theta) = \sum_i f_{i,j}(\phi) \bar{\phi}_i$$

with $\bar{\phi} = \{\bar{\phi}_1, \bar{\phi}_2 \dots\}$ and $\bar{\phi}_i$ is independent of all elements in ϕ in the coefficient of $f_{i,j}$ then the means of $\bar{\phi}$ can be directly substituted. Thus :

$$E_{\bar{\phi}|\phi} [NB(j, \theta)] = \sum_i f_{i,j}(\phi) E(\bar{\phi}_i)$$

and $EVPPI(\phi)$ can be estimated using a single-level Monte Carlo mean.

If $NB(j, \theta)$ can be expressed as:

$$NB(j, \theta) = \sum_i f_{i,j}(\phi) g_{i,j}(\bar{\phi})$$

where each $g_{i,j}(\bar{\phi})$ is a product of mutually independent elements of $\bar{\phi}$, the means of the values in $\bar{\phi}$ can be substituted directly, where:

$$E_{\bar{\phi}|\phi}[NB(j, \theta)] = \sum_i [f_{i,j}(\phi) g_{i,j}(E(\bar{\phi}_1), E(\bar{\phi}_2), \dots, E(\bar{\phi}_i))]$$

and, as in the case above, $EVPPI(\phi)$ can be estimated using a single Monte Carlo mean.

Thirdly, as used in Welton et al. (2008), if $NB(j, \theta)$ can be expressed as [241]:

$$NB(j, \theta) = \sum_i f_{i,j}(\phi) \beta_{i,j}(\bar{\phi})$$

where $\beta_{i,j}(\bar{\phi})$ can be any function of $\bar{\phi}$ (i.e. $\bar{\phi}$ is not necessarily multi-linear), $f_{i,j}(\phi)$ is the linear coefficient of $\beta_{i,j}(\bar{\phi})$ and the elements of $\bar{\phi}$ in $\beta_{i,j}(\bar{\phi})$ are independent of the elements of ϕ in $f_{i,j}(\phi)$ then the NMB can be linearised and parametrised:

$$E_{\bar{\phi}|\phi}[NB(j, \theta)] = \sum_i f_{i,j}(\phi) E_{\bar{\phi}}(\beta_{i,j}(\bar{\phi}))$$

and $EVPPI(\phi)$ can be estimated using a single Monte Carlo mean.

The second set of methods presented involve approximating $E_{\bar{\phi}|\phi}[NB(j, \theta)]$ [128]. Two approaches are presented. If the NMB can be expressed as:

$$NB(j, \theta) = \sum_i f_{i,j}(\phi) \beta_{i,j}(\bar{\phi}_{-i}) h_{i,j}(\bar{\phi}_i, \phi)$$

where $f_{i,j}(\phi)$ are arbitrary functions of ϕ , $h_{i,j}(\bar{\phi}_i, \phi)$ are smooth non-linear functions of ϕ and a single element of $\bar{\phi}$, $\beta_{i,j}(\bar{\phi}_{-i})$ are arbitrary functions of $\bar{\phi}_{-i}$ then the Taylor expansion of $h_{i,j}(\bar{\phi}_i, \phi)$ can be used to approximate $E(\bar{\phi}_i)$.

Spline approximation methods can be used if correlations exist between parameters, so the NMB takes the form:

$$NB(j, \theta) = \sum_i f_{i,j}(\phi) h_{i,j}(\bar{\phi}, \phi)$$

where $f_{i,j}(\phi)$ are functions of ϕ ; $h_{i,j}(\bar{\phi}, \phi_i)$ a smooth non-linear function of $\bar{\phi}$ and at least one ϕ_i is correlated with at least one element of $\bar{\phi}$. A spline can be used to estimate $E_{\bar{\phi}|\phi}(h_{i,j}(\bar{\phi}, \phi))$. This approximation enables the EVPPI to be calculated using a single

Monte Carlo simulation.

All methods presented in Madan et al. (2014) require strong assumptions or special forms for the $NB(j, \theta)$ to give appropriate estimates for the EVPPI [128]. Other methods, which do not require such strong assumptions and enable efficient estimation of the EVPPI are presented below.

Estimation of the expected value of perfect partial information for a single parameter of interest

This section outlines two methods developed for efficiently estimating EVPPI when ϕ is a single parameter, both of which were reviewed in Heath et al. (2017) [95, 187, 212].

Strong and Oakley (2012)

This is a one-level Monte Carlo algorithm requiring a PSA sample of parameters and the corresponding NMBs in its calculation [212]. The basic idea is that a PSA sample of size K is ordered by values of ϕ and split into M subsets of dimension L ($L \ll K, M \times L = K$). The $\phi^{(k)}$ in each subset is assumed to be close to its mean value. Therefore, the remainder of the parameters, $\bar{\phi}$, are an approximate sample from the distribution of $\bar{\phi}|\phi^{(k)}$.

Let $\theta^{(k)} = \{\theta_1^{(k)}, \dots, \theta_S^{(k)}\}$ be the Monte Carlo inputs for the set of S uncertain parameters in the k^{th} PSA sample of the CEA. Let j index the interventions, then $NB(j, \theta^{(k)})$ is the NMB with intervention j evaluated at $\theta^{(k)}$ for a cost-effectiveness threshold λ .

The $NB(j, \theta^{(k)})$'s are ordered by their value of $\phi^{(k)}$. This ordered list is partitioned into M subsets each with L entries ($L \times M = K$). The input parameters are relabelled $\theta^{(l,m)}$ - the l^{th} θ in the m^{th} subset. For each subset and intervention:

$$\hat{\mu}_j^{(m)} = \frac{1}{L} \sum_{l=1}^L NB(j, \theta^{(l,m)})$$

And the EVPPI can be approximated as:

$$\widehat{EVPPI} = \frac{1}{M} \sum_{m=1}^M \max_j \hat{\mu}_j^{(m)} - \max_j \left[\frac{1}{K} \sum_{k=1}^K NB(j, \theta^{(k)}) \right]$$

The main benefit of this method is its ease of calculation. It is more efficient than two-level Monte Carlo estimation and does not require any special form or assumptions for the NMB.

It can be estimated using standard output from a PSA. However, it only works when ϕ is a single parameter. It can also be difficult to choose optimal values for L and M .

Sadatsafavi et al. (2013)

A similar method to Strong and Oakley (2012) was developed by Sadatsafavi et al. (2013) [187, 212]. Like Strong and Oakley (2012), it uses the idea that if a treatment decision is optimal for a value of ϕ it will also be optimal for values close to ϕ [95, 187, 212].

The expectation - maximisation - expectation term (Equation 2.10) is transformed into an expectation - maximisation - maximisation calculation. The PSA output ordered by ϕ is split into $m = 1, \dots, M$ subsets, \mathcal{L}_m , whose sizes depend upon the values of ϕ where the optimal decision changes (i.e. each subset m can have a different size). \mathcal{L}_m contains all ϕ such that $\phi^{\mathcal{L}_m} \leq \phi \leq \phi^{\mathcal{L}_{m+1}}$. In practice, M is normally small. The authors suggest the points where the optimal treatment decision changes can be found by plotting the cumulative NMB as ϕ increases. When the cumulative sum changes direction the optimal treatment decision changes. However, Heath et al. (2017) found this could be challenging in practice [95]. Taking the ordered subsets of NMBs for each m the EVPPI can be estimated (similar to Strong and Oakley (2012)), as [212]:

$$\widehat{NB}^{(m)}(j, \theta) = \frac{1}{L} \sum_{k \in \mathcal{L}_m} NB(j, \phi^{(k)})$$

where L is the size of \mathcal{L}^m . This is maximised over all M subsets:

$$\widehat{EVPPI} = \max_{\phi^{\mathcal{L}_1}, \dots, \phi^{\mathcal{L}_M}} \frac{1}{M} \sum_{i=1}^M \max_j \widehat{NB}^{(m)}(j, \theta) - \max_j \left[\frac{1}{K} \sum_{k=1}^K NB(j, \theta^k) \right]$$

The main benefit of this calculation method is its improved calculation speed compared to the two-level Monte Carlo simulation methods [212]. This method converges in probability to the true value of EVPPI, so it is an asymptotically unbiased estimator. However, it can be sensitive to the choice of M and can only be used to estimate EVPPI for a single parameter. In addition, as with other methods, there is an upwards bias due to the finite PSA sample size.

Estimation of the Expected Value of Perfect Partial Information for multiple parameters of interest

While the methods outlined above are useful and of interest they can only be used when ϕ is a single parameter [187, 212]. The EVPPI for groups of parameters is often more informative for research prioritisation, with any proposed study likely to collect information on multiple parameters.

Three methods for estimating EVPPI for multiple parameters of interest are presented: Heath et al. (2016), Jalal et al. (2015) and Strong et al. (2014) [93, 108, 215]. Two of these methods (Strong et al. (2014) and Heath et al. (2016)) use non-parametric regression to avoid two-level Monte Carlo estimation [93, 215]. As outlined in both papers and summarised below, the purpose of the non-parametric regression is to estimate the first term in Equation 2.10, $E_{\bar{\phi}|\phi} [NB(j, \theta)]$ [93, 215].

For each simulation, $k = 1, \dots, K$, the output from the PSA for the j^{th} intervention can be expressed as:

$$NB(j, \phi^{(k)}, \bar{\phi}^{(k)}) = E_{\bar{\phi}|\phi=\phi^{(k)}} [NB(j, \theta)] + \epsilon^{(k)} \quad (2.14)$$

with $E[\epsilon^{(k)}] = 0$. For each $\phi^{(k)}$ the expectation takes a different value, so the NMB can be thought of as a function of ϕ with unknown form, $g(j, \phi)$:

$$NB(j, \theta^{(k)}) = g(j, \phi^{(k)}) + \epsilon^{(k)} \quad (2.15)$$

Thus, the K sample NMBs from the PSA can be regressed on the K sampled parameter values $(\phi^{(k)})$, to estimate $g(j, \phi)$. Non-parametric methods are used to impose minimal restrictions on the form of g . The two methods of estimating EVPPI use different types of non-parametric regression [93, 215].

The fitted values from the non-parametric regression, $\{\hat{g}(j, \phi^{(1)}), \dots, \hat{g}(j, \phi^{(K)})\}$, calculated from evaluating $\hat{g}(j, \phi)$ at $\{\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(K)}\}$ can be used in the estimation of EVPPI:

$$\widehat{EVPPI}_{\phi} = \frac{1}{K} \sum_{k=1}^K \max_j \hat{g}(j, \phi^{(k)}) - \max_j \frac{1}{K} \sum_{k=1}^K \hat{g}(j, \phi^{(k)}) \quad (2.16)$$

where Monte Carlo estimates over the number of PSA samples are used to estimate the expectations over θ . The second term of Equation 2.16 can easily be calculated by using Monte Carlo estimation. However, estimating as above exploits the positive correlation between the two terms leading to a more precise EVPPI estimate, as explained in Strong et al. (2014) [215].

Strong et al. (2014)

The method in this paper can be used to estimate EVPPI when ϕ has dimension P , $P \geq 1$ [215]. Two alternative non-parametric methods are presented: using a Generalised Additive Model (GAM) or a Gaussian Process (GP) [90, 127].

Generalised Additive Models

In a GAM, $g(j, \phi)$ is a sum of smooth functions for each of the P predictors, $s_p(\phi_p)$, $p = 1, \dots, P$:

$$\begin{aligned} NB(j, \theta) &= g(j, \phi) + \epsilon \\ g(j, \phi) &= s_1(\phi_1) + \dots + s_P(\phi_P) \end{aligned} \quad (2.17)$$

where $E[\epsilon] = 0$. The $s_p(\phi_p)$ are commonly cubic splines, a smooth function represented as a series of piecewise cubic polynomials. Each $s_p(\phi_p)$ can also be expressed as:

$$s_p(\phi_p) = \sum_{l=1}^L \beta_l b_l(\phi_p) \quad (2.18)$$

for a basis of dimension L (equivalent to the number of knots) and basis functions $b_l(\phi_p)$. The basis functions take values over the whole range of ϕ_p . Splines can be made as flexible as wished by altering the number and/or location of the knots/pieces in the spline. The work by Strong et al. (2014) uses the R package `mgcv` which uses the above approach to estimate $g(j, \phi)$ [253]. To estimate $\hat{\beta}_l$, this package by default uses a large number of knots and then uses penalised maximum likelihood estimation with the penalty chosen by cross-validation, to give optimal fit and complexity [253]. Strong et al. (2014) found any number of knots greater than three was sufficient for the purpose of estimating EVPPI [215].

To allow for interactions between the parameters in ϕ which may be present in a health economic model a ‘tensor product’ construction can be used which adds extra terms of the

form $s(\phi_i, \phi_j)$ ($i, j = 1, \dots, P, i \neq j$) to Equation 2.17. These interactions allow for a better fit to the model but they have significant computational cost. Assuming m coefficients which are expected to interact, each with a basis of dimension n , the GAM would need to estimate n^m coefficients. As n^m tends to the PSA sample size, K , the model may not be identifiable. Additionally, increasing K will increase the computational effort required, through both calculating the PSA sample and for estimating the EVPPI. It is for this reason Strong et al. (2014) outline a second non-parametric regression method that enables more efficient calculations when there are multiple interacting parameters of interest [215].

Gaussian Process

As P increases and/or a large number of interactions between the parameters are included, the GAM becomes increasingly impractical. In this case Gaussian Process (GP) regression may be preferable. The basic idea behind a GP regression is the $\{g(j, \phi^{(1)}), \dots, g(j, \phi^{(K)})\}$ are assumed to arise from a multivariate normal distribution with a special mean and covariance structure [127]. Therefore, $g(j, \phi)$ can be expressed as an (arbitrarily) flexible smooth function of its inputs:

$$\{g(j, \phi^{(1)}), \dots, g(j, \phi^{(K)})\} \sim N(H\beta, \sigma^2 \Sigma)$$

where H is a design matrix, β is a vector of the regressors giving the linear relationship between ϕ and the conditional expectation the NMB, Σ is a correlation matrix, taking a squared exponential form in Strong et al. (2014), and σ^2 a constant [215]. Values of H , β and Σ , giving an appropriate level of accuracy and flexibility for the fitted function can be found using numerical optimisation or analytical approaches, as explained in more detail in Strong et al. (2014) [215].

Implementation of Strong et al. (2014)

Both the GAM and GP have reasonable computational cost, akin to calculating the PSA itself. This is still an improvement on the two-level Monte Carlo method. If $P = 1$, these methods can be implemented using the BCEA package in R [15]. This package uses the EVPPI calculation method from Heath et al. (2016) when $P > 1$ as its default (explained below) [93]. However, the BCEA package can be used to estimate the EVPPI using GAM as an option [15]. Alternatively, a web-tool has been developed by Strong et al. (2014), Sheffield Accelerated Value of Information (SAVI), which can carry out these calculations for single and multiple parameters [211]. To use SAVI, the user is required to provide the output from a PSA and the variables to include in the EVPPI calculation (ϕ). The tool calculates the number of ‘knots’,

interactions between parameters, and the choice of GAM or GP for the user - if $P < 5$ GAM is used else GP. However, as the GP method involves inverting a $K \times K$ matrix it uses a maximum of 7,500 PSA simulations. Additionally, R code for the calculations using these methods, in the form of a package SAVI, can be downloaded from the SAVI web-page [211].

To summarise, while GAM is the most straightforward, non-parametric regression approach it is impractical for large P and/or when there are lots of interactions between the parameters. In this case a GP should be used, requiring user interaction to ensure convergence as explained by Strong et al. (2014) [215].

Heath et al. (2016)

Heath et al. (2016) use a specific form of GP regression to estimate multi-parameter EVPPI ($P \geq 2$) [93]. Integrated Nested Laplace Approximation (INLA) is used to fit the high-dimensional GP and is widely used in spatial statistics. It uses the idea that points close to each other, in a geographical way, are thought to have common features and be influenced by common factors more than points that are further away from each other. The EVPPI problem is transformed into a spatial problem by considering the simulated NMBs to be observed at different points in the parameter space [93]. Fast Bayesian computation methods for spatial models can be used. However, spatial statistics are restricted to a two-dimensional space, so the problem needs to be reduced using Principal Fitted Components (PFC).

Computationally, this method has been shown in some cases to be an improvement on the model proposed by Strong et al. (2014) [93, 215]. However, the authors found a small loss in accuracy compared to Strong et al. (2014) when ϕ was of high dimension [93, 215]. This method can be implemented using the BCEA package in R [15]. The main disadvantage of this approach is it cannot be used to estimate EVPPI when $P = 1$.

Jalal et al. (2015)

In addition to the methods by Heath et al. (2016) and Strong et al. (2014) using non-parametric regression, another method has been proposed by Jalal et al. (2015) which can be used for all dimensions of ϕ [93, 108, 215].

This method is similar to the non-parametric methods. However, it assumes $g(j, \theta)$ is linear and that ϕ has a Normal distribution. The relationship between ϕ and the NMB also needs to be approximately Normal. Thus, it would be preferable to use a non-parametric model

approach for $g(j, \boldsymbol{\theta})$ over this method, such as the methods by Strong et al. (2014) and Heath et al. (2016) [93, 108, 216].

2.4 The Expected Value of Sample Information

2.4.1 Background and theory

The EVPPI, introduced in Sections 1.1.3 and 2.3, provides an upper bound for the value of future research for a set of parameters, ϕ , i.e. the value of eliminating all uncertainty on the parameters in ϕ . However, a study collecting further information on ϕ will have a finite sample size whereas an infinite sample size is needed to eliminate uncertainty. The value of conducting a study with a particular design and sample size is presented via the EVSI.

As outlined in Section 1.1.3, let the proposed new study have a sample size of m . Additionally, assume the CEA has j interventions and the parameters of interest for the proposed study be ϕ with the set of all uncertain parameters in the CEA, $\boldsymbol{\theta}$ ($\boldsymbol{\theta} = \{\phi, \bar{\phi}\}$). Let the as yet unknown data generated from the proposed study be \mathbf{X} , with observation x of \mathbf{X} used to gain information on ϕ . \mathbf{X} arises from a statistical model with unknown priors, $\boldsymbol{\theta}$. Therefore, the joint distribution of $\boldsymbol{\theta}$ and \mathbf{X} is:

$$p(\boldsymbol{\theta}, \mathbf{X}) = p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

with the marginal distribution of \mathbf{X} , the posterior predictive distribution being:

$$p(\mathbf{X}) = \int p(\boldsymbol{\theta}, \mathbf{X})d\boldsymbol{\theta}$$

The form of the distribution of $p(\mathbf{X}|\boldsymbol{\theta})$ depends on the proposed study design. The EVSI for \mathbf{X} is:

$$EVSI_{\mathbf{X}}(m) = E_{\mathbf{X}} \left[\max_j E_{\boldsymbol{\theta}|\mathbf{X}} [NB(j, \boldsymbol{\theta})] \right] - \max_j E_{\boldsymbol{\theta}} [NB(j, \boldsymbol{\theta})] \quad (2.19)$$

where $E_{\mathbf{X}}$ is the expectation taken with respect to the posterior predictive distribution of the future data $p(\mathbf{X})$ (i.e. as \mathbf{X} is unknown, the expectation over all potential values of \mathbf{X}). $E_{\boldsymbol{\theta}|\mathbf{X}}$ is the expectation with respect to the updated posterior distribution of $\boldsymbol{\theta}$ given the ‘future data’ has been collected.

The expected benefits from the population-level EVSI can be compared to the expected costs of research. Equation 2.19 presents the EVSI on a per person basis. It can be scaled to a

population-level value using the methods in Section 2.2.2. Should the EVSI be greater than the expected costs there is potential value in carrying out the study.

The Expected Net Benefit of Sampling

EVSI calculations can be used as part of an optimisation exercise which aims to find the optimal sample size for a study by optimising the ENBS over m [2]:

$$ENBS_X(m) = popEVSI_X(m) - Cost(m)$$

where $popEVSI_X(m)$ is the population-level EVSI for a proposed study of size m , reducing uncertainty on parameters ϕ (calculated using the methods of Section 2.2.2 with EVSI used in place of EVPI) and $Cost(m)$ is the cost of a proposed study of size m . The aim of the optimisation exercise is to find:

$$\arg \max_m ENBS_X(m)$$

The costs of a study include direct study costs and opportunity costs. In terms of direct costs, there are fixed costs of running a study such as the cost of a study manager and the cost of equipment. Additionally, marginal costs are associated with each participant in the study, such as device costs, hospital costs, or consultation costs. The financial costs of a study can vary due to their nature and design, such as the length of follow-up.

Opportunity costs of further research are more difficult to quantify. Ades et al. (2004) outline the main types of opportunity costs associated with additional research [2]. Firstly, dependent on the intervention and disease area those involved in the additional research may be unable to benefit from the outcome of this research, meaning a trade-off needs to be made. Increasing the proposed study size would provide more information by reducing uncertainty around the optimal treatment. However, the population who could benefit from this research may be reduced. Similarly, a longer study follow-up delays the presentation of the results reducing the population who can benefit from the study.

While further research is being carried out, those patients not involved in the study often receive the current standard treatment [37]. This may not be the optimal treatment under the CEA due to concerns about irreversibility of treatment decisions. For example, if by implementing a new intervention there is a large sunk cost it is unlikely hospitals would be willing (or financially able) to revert back to the comparator should this be suggested after further research. This means current patients not involved in the proposed study will forgo

the expected additional benefit from the a priori optimal treatment. McKenna and Claxton (2011) outline calculations for costs of delaying implementation for both those in and outside the proposed study [137].

Since repeated EVSI and cost calculations are needed for optimisation, efficient calculation methods for EVSI are useful. As with EVPPI, a number of methods for improving the efficiency of EVSI calculations have been recently published. The focus of this work is on the calculation of EVSI and not the costs and resulting ENBS calculations.

2.4.2 Methods of calculation

This section outlines some recent methods for estimating EVSI. This includes the ‘traditional’ Monte Carlo method and special cases where approximations can be made. A number of developments to improve the efficiency of the calculations are also presented. This review is not meant to be comprehensive, its purpose is to highlight the breadth of recently developed methods. A summary of methods in terms of their populations and their relative merits are presented in Table 2.2.

Monte Carlo methods

Similar to EVPPI, the traditional method for estimating EVSI is via a two-level Monte Carlo simulation. Suppose there are N ‘outer’ and L ‘inner’ simulations, where $N \times L = K$. In the n^{th} simulation, $x^{(n)}$, the data which would be collected by the proposed study is generated by generating a sample, say $\theta^{(n)}$, from $p(\theta)$ (the prior distribution of the parameters θ under current information) and then sampling $x^{(n)}$ from $p(X|\theta = \theta^{(n)})$ where $\theta^{(n)}$ is the n^{th} sample from $p(\theta)$ (the sampling distribution of the data from the proposed new study). The sample prior, $p(\theta^{(n)})$, is combined with the likelihood, $p(x^{(n)}|\theta)$, to obtain the posterior, $p(\theta|x^{(n)})$. Let $\theta^{(n,l)}$ be samples from the posterior distribution of $\theta|x^{(n)}$. If the posterior distribution does not take a closed form then the Monte Carlo EVSI estimator is [216]:

$$\widehat{EVSI}_{\phi} = \frac{1}{N} \sum_{n=1}^N \max_j \frac{1}{L} \sum_{l=1}^L NB(j, \theta^{(n,l)}) - \max_j \frac{1}{K} \sum_{k=1}^K NB(j, \theta^{(k)}) \quad (2.20)$$

$$= \frac{1}{N} \sum_{n=1}^N \left[\max_j \frac{1}{L} \sum_{l=1}^L NB(j, \theta^{(n,l)}) - \frac{1}{L} \sum_{l=1}^L NB(j, \theta^{(n,l)}) \right] \quad (2.21)$$

As with EVPPI, the two-level Monte Carlo method of calculation is computationally intensive. This is one reason why EVSI calculations are rarely carried out, especially when they are

Table 2.2 Reviewed studies for the methods estimating the Expected Value of Sample Information along with their advantages and disadvantages

Method	Summary	Advantages	Disadvantages
Monte Carlo [2, 216]	Two-level Monte Carlo Sampling	Seen as the gold standard method (if computationally feasible) as guaranteed to converge for an infinite sample size	Computationally intensive; EVSI can only be calculated for correlated parameters if they can be expressed in closed form
Ades et al. [2]	(2004) Approximations and simplifications to the two-level Monte Carlo method (when ϕ and $\bar{\phi}$ are independent) using one-level Monte Carlo sampling	Improved computational time compared to the standard two-level Monte Carlo sampling method	ϕ and $\bar{\phi}$ need to be uncorrelated; priors for θ need to be fully parametric; approximations are often required to take advantage of conjugacy
Brennan and Kharoubi (20067) [23]	Laplace approximation to replace Bayesian updating and the inner Monte Carlo integration	Improved computational time compared to the standard two-level Monte Carlo sampling method	Posterior distribution of ϕX needs to be in closed form to take advantage of the calculation efficiencies; distributions of θ need to be smooth, differentiable, and uni-modal
Jalal et al. [108]	(2015) Unit Loss Normal Integral and Linear Regression Meta-modelling	Improved computation time compared to two-level Monte Carlo sampling method; different parameters can have data collected on different sample sizes; R code presented in the paper	prior distributions of ϕ need to be approximately normal; INB and ϕ need to have a linear relationship; difficulties in calculation when there are 3+ interventions
Menzies [138]	(2015) Re-weighting of parameter sets to approximate $E_{\theta X}$ (Importance Sampling)	Improved efficiency compared to the two-level Monte Carlo sampling methods; no restrictions on prior and likelihood distributions; R code available	Trade-off between a simple (yet biased) algorithm and a more accurate (and less biased) algorithm which can be difficult to implement

Method	Summary	Advantages	Disadvantages
Strong et al. (2015) [216]	Non-parametric regression (GAM) to approximate $E_{\theta x}$	Improved calculation time compared to two-level Monte Carlo sampling method; complementary to Strong et al. (2014) EVPPI method [215]; R code available	Can be difficult to generate the sampled datasets and the required summary statistics.
Heath et al. (2017) [95]	Moment matching for estimating $E_{\theta x}$	Computationally efficient compared to the two-level Monte Carlo sampling method; R code available	Less reliable for smaller study sizes and small EVSI values
Jalal and Alarid-Escuero (2018) [107]	Linear regression meta-modelling and Gaussian approximation (extension to Jalal et al. (2015) [108])	Improvement in computational efficiency compared to two-level Monte Carlo sampling method; no distributional form is required for the prior or likelihood; parameters in ϕ can be correlated or interact	Can be difficult to calculate the prior sample size; trade-off between accuracy and over-fitting

required to be performed repeatedly to find optimal sample sizes and/or for different study designs. A further limitation is in obtaining $\theta^{(n,l)}$ for the ‘inner loop’. If $p(\mathbf{X}|\theta)$ and $p(\theta)$ are not conjugate, methods such as MCMC need to be used which further increases the user and computational burden.

Ades et al. (2004)

Ades et al. (2004) simplify the two-level Monte Carlo method for EVSI for special cases [2]. When prior independence exists between ϕ and $\bar{\phi}$ (where ϕ is the set of parameters in θ the proposed study will collect information on) the calculations can be simplified to improve the efficiency of estimating the conditional expectation, $E_{\theta|X}$ (Equation 2.19).

When $NB(j, \theta)$ is linear in ϕ and $\bar{\phi}$ and there is no correlation between any of the elements in ϕ and $\bar{\phi}$, the prior means for $\bar{\phi}$ and the posterior means for ϕ (i.e. the mean after the proposed data has been collected) can be used in the inner-step of the Monte Carlo estimate:

$$E_{\theta|X^{(k)}} [NB(j, \theta)] = NB \left(j, E(\phi|X^{(k)}), E(\bar{\phi}) \right)$$

This can also be used when $NB(j, \theta)$ contains products of independent elements of ϕ and there is no correlation between elements of θ . These are the only situations that fully avoid the inner Monte Carlo simulation.

When $NB(j, \theta)$ is linear in ϕ and non-linear in $\bar{\phi}$, the inner Monte Carlo estimation can be simplified using a nested Monte Carlo integration over $\bar{\phi}$ (i.e. drawing from the prior distributions of $\bar{\phi}$ and using the posterior means for ϕ):

$$E_{\theta|X^{(k)}} [NB(j, \theta)] = E_{\bar{\phi}} \left[NB \left(j, E(\phi|X^{(k)}), \bar{\phi} \right) \right]$$

The following simplifications require the mean and variance of the posterior distributions ($\theta|X$) to be available in closed form. When $NB(j, \theta)$ is non-linear in ϕ and linear in $\bar{\phi}$, a nested Monte Carlo integration over the posterior distribution $\phi|X$ can be used, with the prior means of $\bar{\phi}$ used to obtain an approximate inner Monte Carlo expectation:

$$E_{\theta|X^{(k)}} [NB(j, \theta)] = E_{\phi|X^{(k)}} [NB(j, \phi, E(\bar{\phi}))]$$

If $NB(j, \theta)$ is non-linear in ϕ and $\bar{\phi}$, Monte Carlo integrations need to be carried out using the posterior distribution of $\phi|X$ and the prior distribution of $\bar{\phi}$ and so no simplification of the ‘traditional’ two-level Monte Carlo sampling method is possible.

A major drawback to these methods is that they require no correlation between ϕ and $\bar{\phi}$. Additionally, the priors for θ need to have parametric distributions as conjugacy is required to ensure a closed form.

Brennan and Kharroubi (2007)

Brennan and Kharroubi (2007) presented a form of Laplace approximation to replace the Bayesian updating and the inner Monte Carlo integration step in the two-level Monte Carlo sampling method [23]. This improves the efficiency of the calculation, compared to the two-level Monte Carlo sampling method if the posterior distribution of ϕ can be expressed in closed form. If this is not possible, numerical optimisation methods can be used but these increase the calculation time. For this method to be used, the distributions of θ need to be smooth, differentiable, and uni-modal, meaning the priors cannot have a discrete or empirical distribution.

Menzies (2015)

Menzies (2015) aimed to improve the efficiency of calculating the inner expectation, $E_{\mathbf{X}|\theta}$ [138]. Using the PSA samples from the prior distribution, $p(\theta)$ and Bayes Theorem, a numerical approximation of $p(\theta|\mathbf{X})$ can be obtained by reweighting the parameter sets according to $p(\mathbf{X}|\theta)$. Assuming K PSA samples and values of $p(\theta)$ and $NB(j, \theta)$ for each k :

$$\begin{aligned} E_{\theta|\mathbf{X}}[NB(j, \theta)] &= E_{\theta} \left[\frac{NB(j, \theta)p(\mathbf{X}|\theta)}{p(\mathbf{X})} \right] \\ &\cong \sum_{k=1}^K NB(j, \theta^{(k)}) w_k(\mathbf{X}) \\ \text{where } w_k(\mathbf{X}) &= \frac{p(\mathbf{X}|\theta_k)}{\sum_{k=1}^K p(\mathbf{X}|\theta_k)} \end{aligned}$$

Estimating $E_{\theta|\mathbf{X}}[NB(j, \theta)]$ by re-sampling from the K samples with $w_k(\mathbf{X})$ as the sampling weights is equivalent to Rubin's sample importance i.e. sampling approach for posterior distributions [185]. Two algorithms are presented. The first is simple to implement, but as evidence from the study increases the likelihood is more concentrated in a small region of parameter space leading to many weights close to zero. This reduces the effective sample size leading to over-fitting and upwards bias. This is equivalent to having an inadequate number of 'inner' samples in the two-level Monte Carlo sampling method. The second algorithm is more complex, using smoothing techniques to obtain a better estimate for $E_{\theta|\mathbf{X}}[NB(j, \theta)]$. It

is a generalisation of the EVPPI method by Strong et al. (2014) [215].

The advantage of the Menzies (2015) method is, as with the other approximation methods, it uses PSA output. Additionally, no requirements (such as conjugacy) are required for the prior and likelihood. However, while the first algorithm is simpler to implement it produces biased estimates. The second algorithm, while having minimal bias and increased accuracy, can be difficult to implement in practice due to requiring more user input.

Strong et al. (2015)

Strong et al. (2015) present a method for estimating EVSI similar to the EVPPI calculation method in Strong et al. (2014) (Section 2.3.2) [215, 216]. This uses PSA output and non-parametric regression [215, 216]. Briefly, to calculate the EVSI for collecting information \mathbf{X} using a PSA sample $(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(K)})$ of size K :

1. A data sample, $x^{(k)}$, needs to be generated from $p(\mathbf{X}|\boldsymbol{\theta}^{(k)})$ for each $k = 1, \dots, K$ producing a sample from the posterior predictive distribution.
2. The data from the study, x , can be a scalar or a vector and may provide information on one or more parameter(s), $\boldsymbol{\phi}$. A low-dimensional sufficient statistic, $T(x)$ (which can be one-dimensional or multi-dimensional) is calculated and includes all the information gained from the new study about the parameter(s) $\boldsymbol{\phi}$. A sufficient statistic contains all information required to compute any estimate of the parameter [74]. Formally, $t = T(x)$ is defined to be a sufficient statistic for parameters $\boldsymbol{\phi}$ if the probability distribution of \mathbf{x} , given $t = T(x)$ does not depend on $\boldsymbol{\phi}$ [74]. So:

$$P(x|t, \boldsymbol{\phi}) = P(x|t) \quad (2.22)$$

For example, the sample mean, the sample odds ratio, or parameters of a probability distribution estimated from the proposed study could all be sufficient statistics [216].

3. Using a GAM, the PSA samples $NB(j, \boldsymbol{\theta}^{(k)})$ can be regressed on $T(x^{(k)})$, for each intervention (rather than $\boldsymbol{\phi}^{(k)}$ in the EVPPI calculation) [215]. This calculation method can be simplified by using INB as opposed to NMB in the regression, reducing the number of regressions needed. The fitted values from this regression, $\hat{g}(j, T(x))$, can be used in the EVSI calculation:

$$\widehat{EVSI} = \frac{1}{K} \sum_{k=1}^K \max_j \hat{g}(j, T(x^{(k)})) - \max_j \frac{1}{K} \sum_{k=1}^K \hat{g}(j, T(x^{(k)})) \quad (2.23)$$

As with other methods the PSA output is required, which is readily available. This method is complementary to the Strong et al. (2014) EVPPI estimation, with both using GAM regression [215]. The main drawback is that it can be difficult to generate the sampled datasets and to find the required summary statistic for complex study designs.

Jalal et al. (2015)

Jalal et al. (2015) extend their approach for calculating EVPPI to estimate EVSI [108]. As when calculating EVPPI (Section 2.3.2) a linear regression of the INB on the ϕ from the PSA output. From this, a value for the prior mean and variance of INB explained by ϕ can be calculated.

Assuming the INB is normally distributed, the pre-posterior variance of INB can be estimated. The pre-posterior distribution is the distribution of the posterior mean INB which has been derived from the prior distribution of the INBs and the additional data from the proposed study generated from the prior, i.e. it is the prior distribution of the posterior mean INB before the data is collected. The authors present R code for calculation of the EVSI. This method has the same advantages and disadvantages as the associated EVPPI calculation presented in Section 2.3.2.

Heath et al. (2017)

Heath et al. (2017a, 2017b) outline a method to implement EVSI calculations using moment-matching to estimate the pre-posterior mean, $E_{\mathbf{X}|\theta}$ in Equation 2.19 [92, 94]. It uses standard quantities which can be easily estimated from a PSA: the mean, μ_{θ} , and variance, σ_{θ}^2 , of the $INB(\theta)$. Before the EVSI is estimated, the EVPPI should be estimated to ensure further research on ϕ may be valuable [229]. The fitted values from a non-parametric regression of $INB(\theta)$ on ϕ , $\widehat{INB}_{\phi}^{(k)}$ should therefore be readily available. These can be calculated using the methods of Strong et al. (2014) and Heath et al. (2016), for example [93, 215]. The final quantity required is $Q > 30$ estimates of the variance of the posterior INB after learning \mathbf{X} , computed using Q samples of \mathbf{X} , σ_q^2 , $q = 1, \dots, Q$. The average posterior variance across the Q samples ($\sigma_{\mathbf{X}}^2$) is estimated as the sample mean of σ_q^2 . This value depends on both the variance of the posterior INB (after learning \mathbf{X} estimated by the Q samples) and the prior variance of the INB.

Using methods of moments, the fitted values of the INB can be rescaled:

$$\widehat{INB}_{\phi}^* = \left(\frac{\widehat{INB}_{\phi} - \mu_{\theta}}{\sqrt{\sigma_{\phi}^2}} \right) \sqrt{\sigma_{\theta}^2 - \sigma_{\mathbf{X}}^2} + \mu_{\theta}$$

producing K rescaled INB_{ϕ} values which can be used to estimate the EVSI:

$$\widehat{EVSI} = \frac{1}{K} \sum_{k=1}^K \max \{0, INB_{\phi_k}^*\} - \max \{0, \mu_{\theta}\}$$

This method is computationally efficient. The only parameter that needs to be calculated is the posterior variance for the INB , all other parameters should have been calculated in previous analyses. R code for this method is available [92, 95]. However, the authors found this method is less reliable for smaller proposed study sizes. Additionally, there is a trade-off in the value of Q used, increasing Q provides a more accurate EVSI estimate, but marginally increases computational time. The authors found method is not as reliable as when the EVSI is small, but as EVSI should only be estimated for parameters where EVPPI was sufficiently large, this should not be an issue in practice.

Jalal and Alarid-Escudero (2018)

As with other methods, Jalal and Alarid-Escudero (2018) simplify $E_{\theta|\mathbf{X}}[NB(j, \theta)]$ using the output from a PSA [107]. A Gaussian approximation is used to compute the pre-posterior distribution of $\phi|\mathbf{X}$. The EVSI is re-expressed as the expected opportunity loss from choosing a suboptimal decision and a linear meta-model is used to compute the EVSI given the pre-posterior distribution. However, if the relationship between the loss and ϕ is not linear splines can be used.

The advantages of this method are the computational gains compared to two-level Monte Carlo estimation. Further, no distributional form is directly required for the prior or likelihood; ϕ can contain correlated or interacting parameters. However, it relies on computing a ‘prior sample size’ (n_0) to reflect the strength of the prior information which can be difficult [108].

2.5 The standard error of Expected Value of Perfect Partial Information estimates

2.5.1 The standard error for the Expected Value of Perfect Partial Information when estimated using regression based methods (Strong et al. (2014))

Uncertainty in the estimation of EVPPI when using non-parametric regression can arise from both Monte Carlo error (due to the limited number of PSA samples) and uncertainty in the estimation of the coefficients in the regression model. A method to calculate the error in the EVPPI estimate associated with the estimation of the coefficients in the non-parametric regression, and therefore the GAM, is given in the online Appendix of Strong et al. (2014), and briefly summarised below [215].

The GAM, for an intervention j , can be expressed as a linear parametric model with coefficients β_j and a design matrix \mathbf{X}^* which maps the estimates of the model coefficients $\hat{\beta}_j$ on to the fitted values $\hat{g} = \left\{ \hat{g}(j, \phi^{(1)}), \dots, \hat{g}(j, \phi^{(K)}) \right\}$ i.e.:

$$\hat{g}_j = \mathbf{X}_j^* \hat{\beta}_j$$

The estimated covariance for the sampling distribution of \hat{g} under repeated sampling of a PSA dataset of size K , $g_j|y_j$, where $y_j = NMB(j, \theta)$ can be defined as:

$$\hat{V}_j = \mathbf{X}_j^* V_{\beta_j} \mathbf{X}_j^{*T}$$

where V_{β_j} is the covariance matrix from $\hat{\beta}_j$. The joint distribution for $\hat{\beta}_j$ is asymptotically multivariate normal, so alternative plausible values for $\hat{\beta}_j$, and thus \hat{g}_j can be generated via:

$$g_j|y_j \sim N(\hat{g}_j, \hat{V}_j)$$

For each j , a large number of values (S) of $g_j|y_j$ can be sampled $\tilde{g}_j^{(s)}$, $s = 1, \dots, S$.

Denote the losses for PSA sample k , by:

$$L_{ks} = L(\theta_k, \beta_s) = \max_j [\tilde{g}_j(\theta_k, \beta_s)] - \max_j \left[\frac{1}{K} \sum_k \tilde{g}_j(\theta_k, \beta_s) \right]$$

For each $s = 1, \dots, S$ an alternative plausible value of the EVPPI is:

$$EVPPI_s = \frac{1}{K} \sum_k L(\boldsymbol{\theta}_k, \boldsymbol{\beta}_s)$$

Thus, the standard error in the EVPPI estimate resulting from uncertainty about the GAM coefficients, $se_{\boldsymbol{\beta}}(EVPPI_s)$, can be estimated as the empirical standard deviation of the S values of $EVPPI_s$.

2.5.2 Extension of the Strong et al. (2014) standard error calculation to include Monte Carlo error

Let \hat{E} , the EVPPI estimate with K PSA samples (Equation 2.12) where $\hat{\boldsymbol{\beta}}$ is the estimated model coefficients of the GAM, be expressed as:

$$\hat{E} = \frac{1}{K} \sum_k L(\boldsymbol{\theta}_k, \hat{\boldsymbol{\beta}}) \quad (2.24)$$

Error in this estimate arises not only from uncertainty about $\hat{\boldsymbol{\beta}}$, but also from Monte Carlo error due to the limited number of PSA samples, K . Section 2.2.3 shows the MCSE of the expectation of a random variable based on K Monte Carlo draws is $\frac{s}{\sqrt{K}}$ where s is the standard deviation of the random variable. We want to find the MCSE from \hat{E} , which can be estimated using:

$$\begin{aligned} MCVar(\hat{E}) &= \frac{1}{K^2} \times K \times var(L(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}})) \\ &= \frac{1}{K} var(L(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}})) \end{aligned}$$

where $L(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}})$ is the loss. $var(L(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}))$ can be estimated by Monte Carlo simulation as:

$$\widehat{var}(L(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}})) = \frac{1}{K} \sum_{k=1}^K (L(\boldsymbol{\theta}_k, \hat{\boldsymbol{\beta}}) - \hat{E})^2$$

Therefore, we estimate:

$$MCVar(\hat{E}) = \frac{1}{K^2} \sum_{k=1}^K (L(\boldsymbol{\theta}_k, \hat{\boldsymbol{\beta}}) - \hat{E})^2$$

and thus, the MCSE can be expressed as:

$$se_K(\widehat{EVPPI}) = \sqrt{MCVar(\hat{E})}$$

The two sources of error can be combined to give a total value for the standard error in the EVPPI estimate due to uncertainty in the estimation of the model coefficients in the GAM procedure and the limited number of PSA samples:

$$se(\hat{E}) = \sqrt{se_K^2(\hat{E}) + se_{\beta}^2(\hat{E})}$$

where $se_{\beta}(\hat{E})$ is the standard error of the estimate resulting from the uncertainty about the GAM coefficients (Section 2.5.1). The extra error due to the inclusion of Monte Carlo error can be expressed as:

$$se(\hat{E}) - se_{\beta}(\hat{E})$$

2.6 Exploring the population value of perfect information and their errors using the case study cost-effectiveness analysis

The Monte Carlo error for the EVPI has been calculated and used to determine the appropriate optimal number of PSA samples required for the case study CEA (Section 1.3) for use in the rest of the thesis [198]. Additionally, an estimate of the population-level EVPI for the case study CEA is presented and the impact of uncertainty around the epidemiological quantities required in this calculation have been investigated. The EVPPI for a number of parameters in the case study CEA are presented with standard errors determined by the methods in Section 2.5.

2.6.1 Using Monte Carlo error to calculate the optimal number of simulations for estimating the Expected Value of Perfect Information

To ensure the number of PSA samples is sufficient to enable reasonable conclusions to be drawn, the EVPI for a number of different PSA sample sizes, along with associated errors are presented for the case study CEA (Table 2.3). The expected loss, $E(L)$, with a 95% interval, calculated as $E(L) \pm 1.96 \times s.e$ are shown in Figure 2.2.

As the number of PSA samples increases the MCSE and the width of the 95% interval for the MCSE decreases. Typically, PSAs use 1,000 simulations [91]. On this basis, the NMB and EVPI would be accurate to the nearest £200. The INB between MADs and CPAP in the

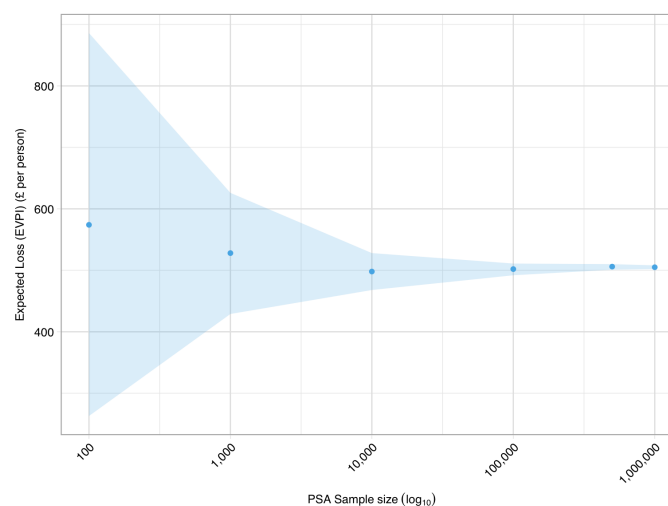
Table 2.3 The Monte Carlo standard error and 95% interval for the expected loss¹ for different sample sizes for the case study cost-effectiveness analysis

Sample Size (K)	E(L) (£)	SD(L) (£)	MCSE (£)	95% Interval (£)	Interval Width ² (£)
100	564	1,590	159.0	(252, 876)	623
1,000	528	1,593	50.4	(429, 627)	198
10,000	498	1,530	15.3	(468, 528)	60
100,000	502	1,547	4.9	(492, 512)	19
500,000	506	1,547	2.2	(502, 510)	9
1,000,000	505	1,546	1.5	(502, 508)	6

¹The expected opportunity loss is equivalent to the EVPI (Section 2.2.1).

² $\approx 4 \times MCSE$

Figure 2.2 The expected loss and its 95% interval for different sample sizes in the case study cost-effectiveness analysis



case study CEA was £281 (Table 1.2) [198]. This result with $K = 1,000$ indicates a high level of uncertainty as to the optimal treatment. Whereas, if $K = 500,000$ the result would be accurate to the nearest £10, indicating a high degree of confidence that CPAP was the optimal treatment.

To judge the EVPI to a suitable degree of accuracy for this analysis, and similar analyses later in the thesis, 100,000 PSA samples appear to be sufficient and gives the EVPI $\pm£10$. This took approximately 90 minutes using four parallel processors. There is a trade-off between the extra computation time and the increase in accuracy from more computation. For example, increasing the number of simulations from 100,000 to 500,000 increases the accuracy of EVPI, from the nearest $\pm£10$ to the nearest $\pm£5$ whilst taking five times longer to compute.

2.6.2 The Expected Value of Perfect Partial Information for uncertain parameters in the case study

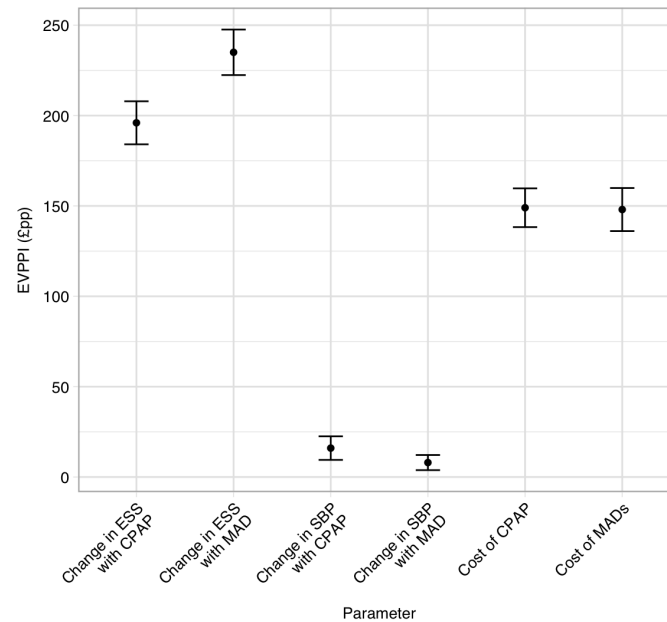
As discussed in Section 2.5, the standard error associated with the EVPPI estimated using non-parametric regression can be calculated. Section 2.6.1 found the number of PSA samples impacts on the accuracy of the EVPI.

The extension of the standard error calculation method in Strong et al. (2014) (Section 2.5) is applied to the case study CEA. Figure 2.3 shows the estimates of the EVPPI for a selection of uncertain parameters in the CEA model whose EVPPI was estimated to be substantially greater than zero. The parameter uncertainty appears to come from a small subset of the parameters.

Figure 2.4 presents the impact the PSA sample size has on the estimates of the EVPPI for the parameters relating to the treatment effect in the case study, the impact of the interventions on the ESS and SBP. As the number of PSA samples increases the EVPPI estimates stabilise.

There appears to be significant value in collecting more information on the impact of both MADs and CPAP on the value of the ESS (Figure 2.4). There is less value in collecting information on the effect of the interventions on SBP. To assess whether these EVPPI values translate into real value of future research, these values need to be transformed into population-level values. The EVSI would also need to be estimated and compared to the

Figure 2.3 The results of the EVPPI estimates for a selection of uncertain parameters in the case study cost-effectiveness analysis¹ using 100,000 PSA samples



¹: The EVPI is £502.

expected costs of future research.

The error associated with the EVPPI estimate for each parameter is presented in Figure 2.5 for a range of PSA sample sizes. The standard error is substantial for PSA samples of less than 10,000. As expected, as the number of PSA samples increases the standard error decreases. The proportion of the error due to Monte Carlo error, even for PSA samples of size 1,000, is small compared to the error due to the GAM approximation alone. This relationship is similar across all sets of parameters.

2.6.3 The population-level Expected Value of Perfect Information for the case study

From Section 2.6.1, using $K = 100,000$ the EVPI estimate for the case study CEA was $£502 \pm 10$. However, this is a per person value. To calculate the upper bound for the value of further research the population-level EVPI needs to be calculated (Section 2.2.2).

OSAHS is often undiagnosed leading to difficulties in estimating its prevalence and incidence. The TOMADO report states 2-7% of the population suffer from OSAHS [173, 198]. Lee et al. (2008) summarised three papers on the population incidence of sleep apnoea, each

Figure 2.4 The results of the EVPPI estimates for parameters in the case study cost-effectiveness analysis relating to treatment effect for a range of PSA simulation sizes

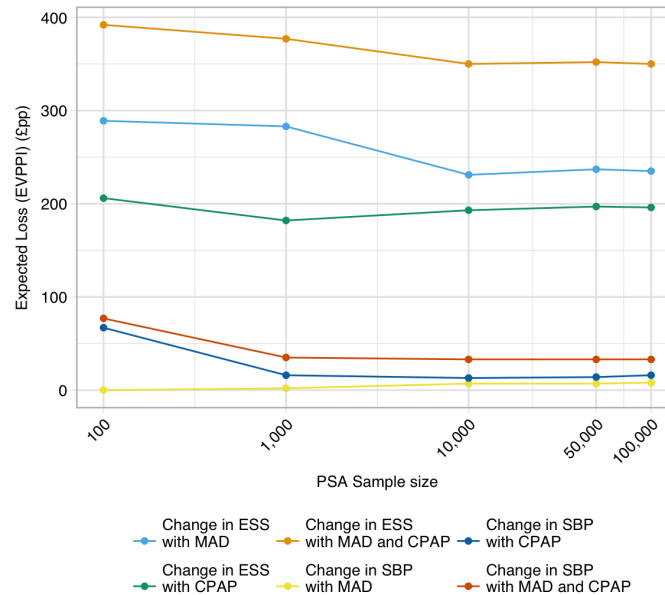
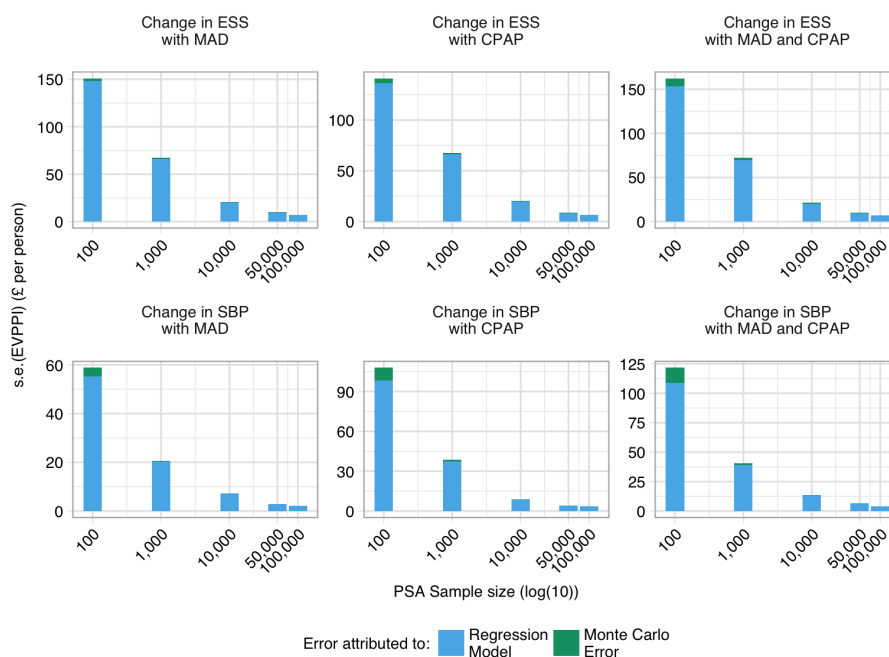


Figure 2.5 The results of the standard error calculation for the EVPPI for various parameters relating to treatment effect in the case study cost-effectiveness analysis incorporating error due to the GAM and Monte Carlo error and presented for a range of PSA simulation sizes



finding the incidence of OSAHS was approximately 2% per annum [122]. However, Young et al. (1997) found over 80% of those with moderate to severe OSAHS and over 90% of those with mild OSAHS were undiagnosed [256].

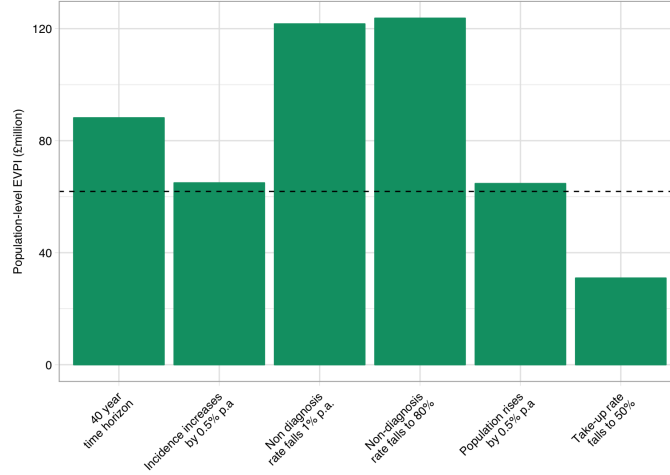
To calculate the population-level EVPI, the population of interest was assumed to be those aged 50-59 in the UK, approximately 3.8 million in mid 2013 (Office for National Statistics mid 2013 year estimates) corresponding to the age of the cohort entering the CEA model [156]. The estimates of incidence (2% per annum) and prevalence (5%) were taken from population based studies [122, 198]. However, only those with diagnosed OSAHS can benefit from further research. An assumption of a diagnosis rate (D_t) of 10% remaining constant over time has been made. In lieu of any disease specific information it has been assumed the take-up rate for the interventions is 100%. A time horizon (T) of 20 years was assumed for MADs and CPAP allowing for future technology development, enhancement, and future improvements in diagnosis rates. The estimated population-level EVPI (popEVPI) (in £ million) is calculated using Equation 2.5 with I_0 the prevalence of OSAHS = 3.8×0.05 , I_t the annual incidence of OSAHS = 0.02×3.8 per annum, i interest rate for discounting = 3.5%, D_t diagnosis rate = 0.1; U_t take-up rate = 1, T time horizon = 20 years, and per person EVPI = 502 (Table 2.3 with 100,000 PSA samples):

$$\begin{aligned} popEVPI &= EVPI \times 3.8 \times \left[0.05 + 0.02 \times \sum_{t=1}^{20} \frac{0.1}{1.035^t} \right] \\ &= £63.76 \text{ million.} \end{aligned}$$

This assumes only 10% of OSAHS cases are diagnosed and those diagnosed take-up their treatment. Figure 2.6 shows the impact of changing the values for time horizon, take-up rate, diagnosis rate, population growth, and prevalence on population EVPI. It is clear that even a modest change in any of the parameters can cause this value to change significantly.

As EVPI and EVSI are both calculated on a per person basis they can be scaled up to population values in the same way. The EVSI, in particular, would need to be compared to the expected costs of research to make the decision on further research. It is important to reflect uncertainty around the population who could benefit from the research. Failure to do this could mean wrong decisions on further research would be made.

Figure 2.6 Population-level EVPI values for a number of different parameters used compared to the baseline population EVPI value^{1,2}



¹ Using the base case assuming a per person EVPI of £506, a prevalence rate of 5%, an incidence rate of 2% per annum, a population of 3.6 million, a discount rate of 3.5%, and a 20 year time horizon

² The horizontal line represents the base case scenario for the population-level EVPI

The impact of the population size on the standard error associated with the Expected Value of Perfect Information

As EVPI values were calculated on a per person basis, the MCSEs associated with the EVPI (Section 2.6.1) are also on a per person basis. It is the MCSE associated with population-level EVPI that directly indicates uncertainty on undertaking further research. Therefore, using the case study CEA and the base case assumptions for population-level EVPI, the impact the number of PSA samples has on population-level EVPI has been explored.

Let P_T be the population set to benefit from the future research, where:

$$P_T = \left[I_0 \times U_0 + \sum_{i=1}^T I_i \times U_i \times (1+i)^{-i} \right] \times P$$

i.e. Equation 2.4 assuming EVPI is £1 per person. Assuming all individuals are independent, the population expected opportunity loss is $P_T \times E_{\theta} [L(\theta)]$. Similarly, the MCSE of the population expected loss, s_p , is:

$$s_p = \sqrt{P_T} s$$

Table 2.4 Monte Carlo standard error and 95% interval for the population expected loss¹ for different numbers of PSA samples for the case study cost-effectiveness analysis

PSA Size (K)	E(L) (£m)	SD(L) (£m)	MCSE (£m)	95% Interval (£m)	Interval Width ² (£m)
100	72	566	56.6	(-39, 183)	222
1,000	67	569	18.0	(32, 103)	71
10,000	63	550	5.5	(52, 74)	22
100,000	64	538	1.7	(61, 67)	7
500,000	64	566	0.8	(62, 66)	3
1,000,000	64	543	0.5	(63, 65)	2

¹The expected opportunity loss is equivalent to the population EVPI.

² $\approx 4 \times MCSE$

where s is the MCSE calculated in Table 2.3. Table 2.4 shows the results of the MCSE for the population of interest. As expected, as the number of PSA samples increases, the width of the 95% interval for the population-level EVPI decreases. However, even with 1,000,000 PSA samples, the value of the population-level EVPI is only known to the nearest £2 million. This figure is large enough to sway the decision of whether further information should be collected. Further, if this value is combined with uncertainty around the epidemiological quantities required to convert from individual to population-level EVPI, where, for example a increase in the diagnosis rate to 20% leads the per person EVPI to approximately double, it is clear the amount of uncertainty on whether future research is viable is large.

2.7 Discussion

Value of information measures can be useful in prioritising further research. However, they are not widely used in practice [207]. This chapter extends on the main value of information quantities - EVPI, EVPPI and EVSI introduced in Section 1.1.3. A number of recent developments in calculation methods have been described, all of which aim to increase the speed of the calculation of EVPPI and EVSI. This chapter has also shown the importance of an adequate number of PSA samples. An insufficient number of PSA samples can lead to errors large enough to alter the decision about the optimal treatment and whether further information should be collected.

Estimating the size of the population who could benefit from additional research has highlighted the importance of presenting population-level estimates of value of information quantities and of acknowledging uncertainty. For diseases with a large prevalence, such as OSAHS, modest per person values can translate to substantial population-level values. When scaling values up to a population-level, it is important to correctly estimate the time horizon, diagnosis rate, and uptake of the interventions and acknowledge their uncertainty. Assuming 100% uptake and full diagnosis can lead to significant over-estimation of the population-level EVPI.

2.7.1 Limitations of the current methods

Value of information quantities were not found to be widely used in practice due to various challenges in computation and interpretation [20]. Bindels et al. (2016) carried out focus groups in Holland to find out what researchers, policy makers, and pharmaceutical companies thought about value of information and their perceived barriers to its use [20]. All participants agreed value of information quantities were useful to guide further research. However, the main barriers to implementation were that: EVSI may suggest infeasible research designs are optimal, it is not easy to incorporate all types of structural uncertainties in the estimates, the quantities can be complex to calculate, and policy makers can have a limited knowledge on the interpretation of the results. Bindels et al. (2016) recommended guidelines should be developed to assist researchers [20].

Despite the wealth of new literature on calculating EVPPI and EVSI quantities, not all methods are simple to understand or calculate. A number of the methods outlined can be implemented using ready-made R packages or Excel workbooks. The BCEA package can calculate the EVPPI for a single parameter using Strong et al. (2014) and for multiple parameters using Heath et al. (2016) [15, 93, 215]. Similarly, the Strong et al. (2014) method can be calculated using an R package, SAVI, and a web-based app [211, 216]. Wilson (2014) provided access to Excel workbooks to help with analytical calculations of EVPI, EVPPI and EVSI [251]. At present, there is little by the way of readily available code for calculating EVSI, however Heath et al. (2017) are in the process of developing this for their work [95]. It is hoped these interfaces and easily accessible functions will lead to an increase in the application of value of information methods. However, it is important and advisable that these methods are used by practitioners who understand their limitations. This is fundamental for the transparency of calculations provided to policy makers [240].

Many of the EVSI calculation methods presented are very recent, so little work compares them. As with EVPPI, all new EVSI methodologies were published alongside comparisons to the two-level Monte Carlo method and all found their methods performed comparably in terms of the size of the estimates but with improvements in the speed of calculations. A study comparing the EVSI estimation methods to one another, as for EVPPI in Heath et al. (2016), would be useful to help guide practitioners on the best method to use [93]. It is understood that this is in progress.

A full value of information analysis is not always required [229]. The EVPI, which is easy to calculate non-parametrically, should be calculated routinely. If this and its associated population-level EVPI are low we can be confident there is little value in further research. In this case calculating EVPPI and EVSI would serve little purpose other than adding to the computational burden.

Conversely, should the EVPI indicate a substantial upper bound for the value of future research, researchers should calculate the EVPPI for the (groups of) parameter(s) of interest. Knowledge of the disease area and potential study designs should be used to calculate EVPPI for clinically relevant parameter(s) which are feasible to collect in a single study, as opposed to data dredging for potential study designs. This will ensure the calculation burden of EVPPI is restricted to information that could be collected in practice. The EVPPI provides an upper bound for the elimination of uncertainty so can be used to guide if and for which parameters the EVSI should be calculated. Only if the EVPPI was found to be substantial should the EVSI be calculated. It is this form of incremental analysis Heath et al. (2017) use to improve the efficiency of their EVSI calculation method [95]. Implementing value of information analyses in this hierarchical manner ensures researchers' time is appropriately used and prevents calculations with no value from being carried out.

The value of information methods presented provide the value of resolving uncertainty in the parameters, θ , from the CEA model. They do not, in their current form, quantify the value of eliminating structural uncertainty. Should there be questions around the uncertainty in the model structure this would not be reflected. As an example, in the case study CEA, a causal relationship is assumed between the impact of treatment and the CVD risk. However, the exact nature of this relationship is unknown and this relationship is modelled through the impact of treatment on SBP. Should further causal mechanisms be identified, the structure of the CEA model may need to be updated. Value of information measures would not be able to quantify the value of collecting information on this causal relationship without updating the

model structure.

Some theoretical literature is available on incorporating structural uncertainty in value of information calculations. Jackson et al. (2011) showed structural uncertainty can be parametrised using elicitation of weakly informed parameters or model averaging [106]. Strong and Oakley (2014a) presented a method which allows for structural uncertainty to be parametrised in a CEA by adding extra parameters, known as discrepancy parameters, on which value of information quantities can be estimated [213]. Price et al. (2011) found through a model averaging approach that EVPI can be sensitive to structural assumptions of the CEA [172]. Despite these methods being present they are not widely used, with structural uncertainty often ignored.

The case study CEA (Section 1.3) does not allow for value of information quantities to be calculated for quantities currently expressed as point estimates, such as adherence to interventions. Sharples et al. (2014) acknowledged that adherence to interventions is uncertain, but due to a lack of data point estimates were used. This appears to be a contradiction, as including a value as a point estimate is the same as stating the value is known with certainty. Chapter 3 explores modelling adherence and calculating the value of collecting further information on adherence to interventions. Similarly, the value of collecting further information on heterogeneity in the CEA population cannot be quantified using the current case study CEA and value of information methods. Chapters 4 and 5 expand the case study CEA model and value of information methods to enable these calculations.

2.7.2 Future research priorities

In recent years, a wealth of literature has been published to improve the speed of calculating value of information quantities. However, as outlined in Welton and Thom (2015) being able to calculate these quantities efficiently is not sufficient to ensure the calculation methods are applied [240]. As mentioned previously, the current methods do not take into account structural uncertainty within the CEA. Further research on how the quantities could be estimated for structural uncertainty would be useful. This would help give a rounded view on the value of collecting further information for all areas of uncertainty.

As mentioned in Bindels et al. (2016) and Welton and Thom (2015), a barrier to value of information being routinely presented is whether policy makers have sufficient knowledge to interpret them [20, 240]. Therefore, education of policy makers and guidelines as to when value of information quantities should be calculated would assist in this. For example,

thresholds in terms of population-level EVPI and EVPPI to indicate further information could be sought.

For EVSI to be used in practice a comparison to the costs of the research needs to be carried out. Current methodological work has focussed on the estimation of EVSI. Methods to estimate the costs of research and the impact of delaying treatment decisions have been neglected. These form an important role in deciding future research priorities and should receive further attention. Brennan and Kharroubhi (2007) note costs and health benefits of a study may be broader than those included in the EVPPI and EVSI calculations so the estimated quantities may be under-estimated [23]. For example, a study of a different population to the target population could still be useful, perhaps through inclusion in a meta-analysis.

Calculating the standard errors for the Strong et al. (2014, 2015) EVPPI estimates in the case study has shown the importance of the PSA sample size [215, 216]. PSA sample sizes are often chosen arbitrarily with many HTA submissions using 1,000 PSA samples [91]. NICE guidelines state models should be run until convergence with no formal definition of convergence defined [150]. Further work on what is an appropriate level of accuracy would help researchers ensure their results are sufficiently accurate and prevent incorrect treatment decisions being made. As a starting point, we recommend researchers present their results rounded to their degree of accuracy. It has been shown (Sections 2.2.3 and 2.5) that the MCSE for the EVPI and the EVPPI using non-parametric regression methods are simple to calculate. Presenting either the error associated with the PSA sample size or rounding the results to the appropriate degree of accuracy would be a simple way to present the uncertainty. A large amount of uncertainty is less problematic if the CEA indicates an intervention is very likely/unlikely to be cost-effective. For example, should an ICER be around £250,000 per QALY gained to the nearest £10,000 the decision around cost-effectiveness is unlikely to change with further PSA samples. However, should an ICER be £20,000 per QALY gained to the same degree of accuracy, further PSA samples could alter the optimal treatment decision.

Section 2.6.3 showed the population-level EVPI is sensitive to quantities such as incidence, prevalence, time horizon, uptake, and diagnosis rates used in scaling the per person EVPIs to a population-level. This can lead to incorrect decisions being made on undertaking future research. Grimm et al. (2015) found many studies did not present population-level EVPI and those that did rarely presented uncertainty around this value [86]. A probabilistic approach to calculating population-level EVPI would be a useful addition to the researchers' toolkit. It would enable distributions to be placed around those parameters where there is uncertainty,

and enable a range of population-level EVPIs to be calculated. Additionally, the value of collecting more information on these parameters could be estimated.

2.7.3 Value of information in the rest of the thesis

One of the aims of this thesis is to identify areas of uncertainty where their reduction would be of particular use (Section 1.4). Therefore, applications of the value of information methods presented in this chapter are used throughout the thesis applied to the case study CEA (Section 1.3 and Appendix B) [198].

Throughout this thesis, the Strong et al. (2014) EVPPI and the Strong et al. (2015) EVSI methods are used [215, 216]. These methods have been chosen for a number of reasons. For the EVPPI, this method can be implemented for both single and multiple parameters of interest. They are simple to implement with access to the R code underlying these methods available enabling easier adaptation to the situations considered throughout the thesis. As outlined in Section 2.5, the standard error due to the uncertainty in the coefficients of the non-parametric regression model and the limited number of PSA samples can be calculated. As the methods for EVPPI and EVSI published by Strong et al. (2014, 2015) are complementary, both using GAM non-parametric regression, concise code could be used for the EVSI method [215, 216].

A PSA sample size of 100,000 is used in the application of the case study CEA in the rest of this thesis. This gives EVPI to $\pm£10$ per person. However, the population-level EVPI is given to the nearest £7 million, which is still vague. This PSA sample size was chosen as a trade-off between the accuracy and the computational time needed to run the CEA. However, increasing the sample size will not help in the context of potentially greater uncertainties about factors governing the population-level EVPI.

While the ENBS is an important concept this will not be considered in further in this thesis. The focus is on how value of information quantities can be calculated to prioritise which parameters should be investigated further and on the ranking different research designs.

2.8 Conclusion

This chapter presented the different value of information quantities available to estimate parameter uncertainty and prioritise future research. A number of recent methods for efficient calculation of value of information quantities have been summarised. Through application of

EVPI and EVPPI methods to the case study CEA an appropriate sample size for the PSA has been determined to provide results to a suitable degree of accuracy for this thesis. Additionally, through an application of EVPPI methods to the case study CEA those parameters where the collection of future information would be valuable have been identified.

The methods presented are used in Chapter 3 to prioritise further data collection in the context of modelling adherence to interventions in a CEA. The EVPPI and EVSI estimation methods by Strong et al. (2014, 2015) are used in Chapter 5 to estimate the value of further research on factors relating to stratification of the optimal treatment decision and the impact of study design on the EVSI for the same set of parameters [215, 216].

Chapter 3

Modelling adherence to interventions in a cost-effectiveness analysis

The proportion of individuals using their intervention as prescribed is the adherence rate. Whether an individual adheres to their intervention impacts their treatment effect. Rates of adherence can change over time. This is not frequently modelled in a CEA. This chapter looks at modelling adherence to interventions applied to the case study CEA through a Bayesian time-to-event meta-analysis of observed adherence data [198]. The impact on the CEA of modelling adherence compared to using point estimates of the adherence rates is examined. Additionally, drawing on the methods from Chapter 2 the value of collecting further information on adherence and the type and size of the study most preferable to conduct is estimated to help guide further work.

3.1 Introduction

All cost-effectiveness models are subject to uncertainty [84]. This chapter looks at one under explored area of uncertainty in the inputs to a CEA: patients adherence¹² to an intervention using the case study. Sharples et al. (2014), the case study CEA, stated under implications for research priorities, *"Similarities of effects for CPAP and MADs on EDS may be due to differential adherence to treatment. However, there is limited information on this beyond*

¹Adherence is how well an individual follows a recommendation having agreed to it, suggesting the individual takes an active role in their treatment whereas compliance is seen as more passive, referring to usage since prescription date [223, 235].

²This work does not take into account the difference between intentional and non-intentional non-adherence. Non-intentional non-adherence occurs when a patient does not adhere due to factors beyond their control (such as difficulties in understanding instructions or inability to pay. Intentional non-adherence occurs when the patient makes an explicit decision to not use prescribed treatment

short term trials. Medium to long-term compliance with MAD and CPAP should be monitored and reported." indicating adherence to these interventions is important to consider [198].

In Sharples et al. (2014) literature on adherence to MADs and CPAP was searched to find values for adherence to the interventions to populate the CEA [198]. Data from a single study, Kohler et al. (2010), provided information on adherence to CPAP [119]. No comparable data was found for adherence to MADs. Therefore, the same point estimates of adherence were used for both interventions [198]. This is a strong assumption which is investigated in detail throughout this chapter.

As a sensitivity analysis, a one-way conservative adjustment on adherence to CPAP was made to take into account that adherence may differ by severity [198]. When adherence to CPAP was reduced by 5%, the ICER between treatment with MAD and CPAP was £40,668 per QALY gained compared to £15,467 per QALY gained in the base-case [198]. This suggests that, at thresholds used by NICE, CPAP is no longer cost-effective [150]. When adherence to CPAP was reduced by 10% treatment with CPAP was dominated by MADs (MADs were more effective and less costly than CPAP) [198]. This simple, deterministic, scenario analysis does not formally quantify uncertainty around adherence to MAD and CPAP. Additionally, no justification appears to be presented as to why adherence to CPAP was reduced as opposed to adherence to MADs.

Perpetual adherence is often assumed in CEAs and rarely questioned except implicitly through sensitivity analysis on the treatment effect [104]. Many CEA models use data from RCTs or observational studies to estimate quantities related to the effectiveness of the treatment. However, these studies are often short term in nature whereas model-based CEAs often take a lifetime perspective, especially for chronic diseases. Participants in RCTs or similar studies may not be representative of real practice - individuals in clinical trials are often self-selecting and subject to more scrutiny than the general population [76]. For usage to be replicated when the drug is licensed for use in health services, it is important that real practice is reflected. If a CEA model does not adequately reflect true adherence to interventions a suboptimal treatment may be recommended for implementation, leading to costs to the population both monetary and in terms of health foregone.

Hughes et al. (2007) identified a number of studies that incorporated measures of adherence in their pharmacoeconomic evaluations [104]. Of the ten included studies only one measured the probability of adherence through the use of a Bernoulli random variable [104]. This

random variable was generated to indicate adherence by an individual independent of effectiveness in each cycle. The probability of adherence at the end of the cycle was pre-defined [42, 104]. Additionally, Brilleman et al. (2016) reviewed trial based economic evaluations to identify how adherence has been reported in trials and its impact on economic evaluations [26]. They found no study adjusted for non-adherence directly in their economic evaluation [26].

This chapter explores adherence to interventions. There are two main aims. Firstly, a method of modelling the change in adherence over time to interventions, using Bayesian meta-analysis of all available data, is introduced as an alternative to using point estimates [198]. Bayesian methods allow uncertainty about adherence rates to be more fully quantified and propagated to the CEA results. This method is applied to the case study CEA (Section 1.3 and Appendix B) to examine the impact on cost-effectiveness of modelling adherence in this way [198]. The second aim of this chapter is to explore the value of collecting further information on adherence to interventions in particular, at what time horizons adherence data should be collected for most benefit, again, applied to the case study CEA.

The chapter is set out as follows: Firstly the methodology and results for the literature review used as the basis of the meta-analysis are reported (Section 3.2). Section 3.3 explains the methodology of the meta-analysis; the incorporation of the meta-analysis results into the CEA and the methods used to assess the value of future research on adherence. Section 3.4 presents the results of the meta-analysis; the CEA and the value of future information calculations. Section 3.5 discusses the findings including limitations of the data, methodological issues and future research priorities before Section 3.6 summarises the conclusions of the chapter.

3.1.1 Current guidance on including adherence to interventions in cost-effectiveness analyses

NICE Methods Guidance issued in 2013, used in England and Wales, defines adherence as *"the extent to which a person follows the health advice agreed with healthcare professionals"* [150]. It mentions *"if characteristics of healthcare technologies have a value to people independent of any different effect on health, the nature of these characteristics should be clearly explained and if possible the value of the additional benefit should be quantified"* [150].

The Canadian Agency for Drugs and Technologies in Health (CADTH) 3rd edition of guidance (2006) defines adherence in three parts: the acceptance of the treatment by the patient; persistence - the long-term continuation of the treatment; and the consistency and accuracy of following the recommended treatment regimen [33]. The guidance states the reference case should incorporate real world factors which may modify the effect of the intervention and adverse events which may impact adherence. It notes adherence may differ in a 'real world' setting, leading to lower treatment effects, higher costs, decreased productivity, and a greater burden on care givers and increased drug resistance [33].

In Australia, the Pharmaceutical Benefits Advisory Committee (PBAC) guidelines expect compliance to be modelled in terms of its impact on improving health outcomes or reduced provision of other healthcare resources [162].

The aforementioned guidance for CEAs indicate that modelling adherence (or compliance) to interventions is important. However, there appears to be no specific guidelines on how adherence should be modelled in practice.

In addition, Hughes et al. (2007) in their work as part of the ISPOR Medication Compliance and Persistence Special Interest group conclude that compliance should be an integral part of pharmacoeconomic evaluations [104]. However, they note available methodology is sparse and limited.

3.2 Literature review

Sharples et al. (2014) performed a literature review of studies of patients with OSAHS treated with MADs or CPAP reporting information on adherence to the intervention [198]. The results of this review were used to provide information to populate the CEA.

Following their review, data from Kohler et al. (2010) was used as the measure of adherence to the intervention in the case study CEA [119, 198]. This was chosen as it was a large hospital record based study of 600 patients' adherence to CPAP. It provided ten years of data and in terms of AHI the population was similar to that in the CEA. However, the population in Kohler et al. (2010) had a slightly higher mean ESS: the case study CEA population had a baseline ESS of 11.9 compared to a baseline ESS of 15.0 in Kohler et al. (2010) [119, 198]. As little data was available on adherence to MAD and there was no evidence that adherence to MAD and CPAP would differ, Sharples et al. (2014) used the same data for adherence to

MAD [198]. The data from this review is a useful starting point to assess the availability of data on adherence to MADs and CPAP as treatment for patients with OSAHS.

For this thesis, the literature review was updated to extract all available data (up to January 2015) on adherence to MADs and CPAP including any new studies published since Sharples et al. (2014) [198].

3.2.1 Search criteria and methodology

Medline was searched in January 2015 using the search terms in Appendix C. Figure 3.1 presents the results of the search and the classification process.

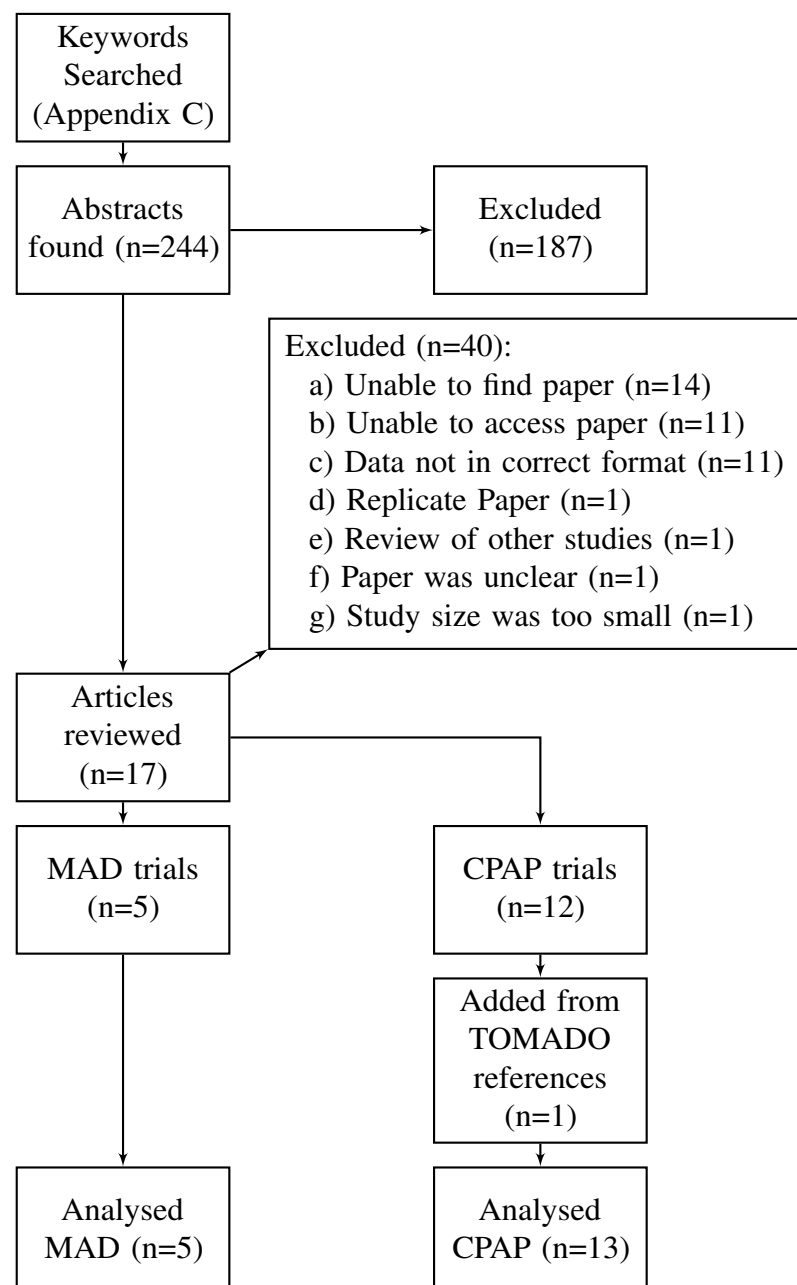
244 abstracts were found and screened by title and abstract. Studies were considered relevant if they included MAD or CPAP as treatments for OSAHS, had at least one year mean follow-up, and reported a measure of adherence over time. Studies were limited to those with more than fifty patients and were written in English. Those studies reporting average usage (in minutes) per night were excluded as it was difficult to transform this into a binary adherence measure ($n=3$) [36, 168, 217].

Fifty papers were reviewed in more detail. The reasons for inclusion or exclusion are given in Appendix D. Of these 50 papers, 17 were included in the final review. The main reasons for exclusions were: the abstracts were conference abstracts with no related full paper (14 papers), inability to access the data (11 papers), or the data was not in the correct format, for example presented the work as hours usage a night (11 papers). Other reasons for exclusion were: the paper was unclear; the paper was a replicate of another paper, the paper was a review of other studies, or the population was too small for inclusion (one paper each). An additional paper was included which was not found in the literature review. This was the study used by McDaid et al. (2009) for their adherence parameters (McArdle et al. (1999)) [134, 135]. One included paper was a conference abstract, Quinell 2014 [174]. This was included as it gave information on adherence to MADs in the TOMADO population, the RCT from Sharples et al. (2014) [198].

3.2.2 Results

Eighteen papers were included in the final review - five looking at adherence to MADs [82, 112, 132, 159, 174] and 13 for adherence to CPAP [4, 28, 31, 34, 77, 79, 115, 119, 121, 135, 228, 233, 236]. Final follow-up times for the papers ranged from one to ten years.

Figure 3.1 Flow diagram of the selection process of the literature search for papers on adherence to MAD and CPAP as interventions for patients with OSAHS



Adherence was measured at timepoints ranging from one month to ten years with each study reporting the proportion of individuals adherent at between one and ten timepoints (Table 3.2). The data is limited for adherence to MADs. For adherence to CPAP there is more information on shorter term adherence (at less than a years usage), but over a longer term the data become sparse with only two studies providing information on CPAP past five years (Tokunaga et al. (2013) and Kohler et al. (2010)) [119, 228].

A number of different definitions of adherence were used in the studies (Table 3.1). In this work, for those studies which presented adherence using multiple definitions the results for the definition closest to more than four hours a night on 70% of the nights has been used. This definition is used in four studies and defined in Shapiro and Shapiro (2010) [4, 77, 121, 195, 233].

The RCTs in the meta-analysis of treatment effects carried out as part of Sharples et al. (2014, 2015) were all short term in nature with follow-up times of less than one year. Therefore, their results and information on adherence could not be used in this review [198, 199].

The average age of the populations across the studies reporting adherence to MADs ranged from 50 years (Pancer et al. (1999)) to 58.5 years (Ghazal et al. (2009)) (a mean age of 50.5 years was used in the case study CEA) [82, 159]. Similarly, the BMI (range: 25.9kgm^{-2} (Ghazal et al. (2009)) to 30kgm^{-2} (Jauhar et al. (2008) and Pancer et al. (1999))) and proportion of males (range: 77.8% (Jauhar et al. (2008)) to 86% (Pancer et al. (1999))) appear consistent between the studies reporting adherence to MADs [82, 112, 159]. The AHI varied between the studies from a low of 5.4 events per hour (Ghazal (2009)) to a high of 37 events per hour (Pancer (1999)) meaning, based on the sleep apnoea severity definition (Department of Sleep Studies, Harvard) the range of patient severities encompassed patients with mild and severe OSAHS [5, 60, 82, 159]. For reference, the TOMADO population had an average AHI of 13.8 events per hour at baseline (classified as mild OSAHS)³ [5, 60, 198].

There is evidence of heterogeneity across the studies exploring adherence to CPAP. Not all papers looking at adherence to CPAP reported all the summary statistics, so the values quoted are based on available data. The proportion of males varies from 0% (Campos-Rodriguez et al. (2013)) to 100% (Van Zeller et al. (2013)) [32, 233]. Excluding Campos-Rodriguez et al. (2013), all studies had over 78% males (the case study CEA model has a cohort of males) [32]. The average age across studies ranges from an average of 48.7 years (Waldhorn et al.

³This was the TOMADO population, the CEA does not make use of AHI

Table 3.1 Definitions of adherence to interventions for patients with OSAHS in the studies identified by the literature review¹

Intervention	Paper	Definition of adherence
MAD	Ghazal (2009) [82]	More than five nights a week
	Jauhar (2008) [112]	Daily; Usage up to six nights a week
CPAP	Marklund (2006) [132]	Daily; > 50% of the time
	Pancer (1999) [159]	Daily
	Alves (2012) [4]	Any usage; > 4hrs for > 70% days
	Bromstom (2007) [28]	> 50% self reported sleep time; > 4hrs a night
	Campos-Rodriguez (2013) [32]	Any
	Chai-Coetzer (2013) [35]	> 4hrs a night
	Furukawa (2014) [77]	> 4hrs for > 70% of days; < 4hrs for > 70% of days
	Johnson (2004) [115]	Any; > 2hrs a night
	La Piana (2011) [121]	> 4hrs for > 70% days
	van Zeller (2013) [233]	Any ; > 4hrs a night for > 70% of days

¹: For studies that reported a definition of adherence

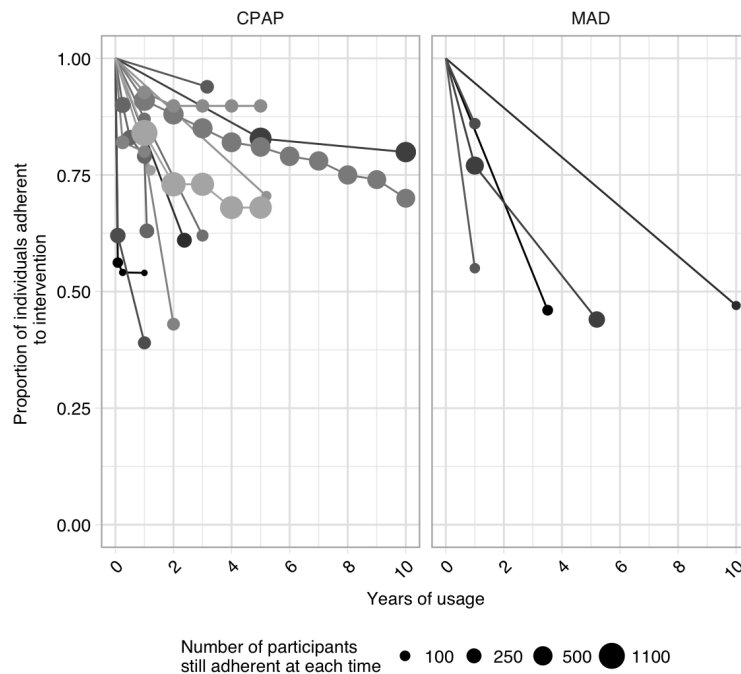
(1990)) to 60 years (Campos-Rodriguez et al. (2013))(the mean age in the case study CEA model was 50.5 years) [32, 236]. The AHI varied from 36.02 events per hour at baseline in Tokunaga et al. (2013) to 57.20 events per hour at baseline in Waldhorn et al. (1990), meaning all studies had a population with severe OSAHS [228, 236]. For reference, the TOMADO population had a mean AHI of 13.8 events per hour at baseline⁴ representing mild OSAHS [198]. The BMI varied from 26.4kgm^{-2} (Tokunaga et al. (2013) and Furukawa et al (2014)) to 36kgm^{-2} (La Piana et al. (2011)) (31.9kgm^{-2} in the case study CEA) [77, 121, 228].

A number of papers explored factors that may affect adherence rates [4, 28, 32, 34, 79, 82, 115, 119, 121, 132, 135, 228, 233, 236]. These can be split into two categories - baseline characteristics of the population and initial usage characteristics. For baseline characteristics (age, AHI or Obstructive Hypopnoea Index (OHI), BMI, and ESS) only some studies found them to be significant in predicting adherence [28, 32, 34, 79, 119, 132, 135, 228, 233, 236]. Those looking at initial usage characteristics found high side effects and low initial adherence led to a decrease in future adherence [115, 135, 233]. Due to a lack of data for each specific factor and uncertainty on which factors impacted adherence, it was decided to note these predictors of adherence but not include them in the modelling framework. All types of MADs and CPAP are grouped together due to a lack of data on individual device types. This could be another cause of heterogeneity between studies. Although there is insufficient data to test this it is worth considering. Borel et al. (2013) found the type of mask used with the CPAP machine may impact adherence [21].

Figure 3.2 shows the adherence rates to MAD and CPAP as interventions for OSAHS found through the literature review. Each shade of grey represents a different study with the size of each point proportional to the number of patients adherent at that timepoint in the study. The lines connecting points from the same study are for ease of interpretation only. There may not be a linear decrease in adherence in practice between two timepoints. Heterogeneity between studies is clear and there is evidence of a high initial hazard of non-adherence, especially to CPAP. The evidence from the literature review indicates adherence to CPAP appears to be different to adherence to MADs. There is some evidence that increased adherence to CPAP may be a result of the study population having more severe OSAHS. To assess the impact of disease severity on adherence more information would be needed on adherence to CPAP in less severely affected populations. No study presented information on adherence beyond ten years usage. Information on longer-term adherence would be useful to explore adherence

⁴As before, the CEA does not use the AHI severity measure

Figure 3.2 Observed adherence data¹ on long-term adherence to MADs and CPAP as treatments for OSAHS taken from the studies analysed as part of the literature review²



¹ The adherence data is taken from the proportion of participants eligible at each timepoint (i.e. excludes those lost to follow-up)

² Each shade of grey represents a different study. The size of the point is proportional to the number still adherent at each timepoint. The lines connecting points from the same study are for ease of interpretation only. The proportion adherent between timepoints may not be linear in practice.

over a patient's lifetime.

3.3 Methods for modelling adherence

The original case study CEA highlighted evidence that the optimal treatment decision may be sensitive to assumptions on adherence to the intervention (Section 1.3) [198]. Additionally, the literature search found little information on adherence to MADs and for the population of interest using CPAP (Section 3.2.2). This section presents a method for using all available data to model the uncertainty around adherence to MADs and CPAP. A fully Bayesian time-to-event meta-analysis model on the adherence data is used to capture the uncertainty in adherence rates and provide estimates for use in a CEA to assess the impact of this uncertainty on the optimal treatment decision.

Table 3.2 Data summary for the proportion of participants adherent to interventions at measured times taken from included papers in the literature review

Reference	Size	Time on intervention (Months)												
		1	3	6	12-23	24-35	36-47	48-59	60-71	72-83	84-95	96-107	108-119	120
MAD														
Gahazal (2010) [82]	103						0.46							
Jauhaur (2008) [112]	72													0.47
Marklund (2006) [132]	423				0.77			0.44						
Pancer (1999) [159]	121				0.86									
Quinnell (2014) [174]	96				0.55									
CPAP														
Alves (2012) [4]	96	0.56	0.54		0.54									
Bromstrom (2007) [28]	247				0.61									
Campos-Rodrigues (2013) [32]	708													
Chai-Coetzer (2013) [34]	275	0.62			0.39			0.83				0.80		
Furukawa (2014) [77]	194						0.39							
Galetke (2011) [79]	303		0.9	0.83	0.79	0.63								
Johnson (2004) [115]	158				0.87		0.62							
Kohler (2010) [119]	639				0.91	0.88	0.85	0.82	0.81	0.79	0.78	0.75	0.74	0.70
La Piana (2011) [121]	202		0.82		0.80	0.43								
McArdle (1999) [135]	204				0.93	0.90	0.90	0.90	0.90					
Tokunaga (2013) [228]	88								0.71					
Van Zeller (2013) [233]	96				0.76									
Waldhorn (1990) [236]	1,103				0.84	0.73	0.73	0.68	0.68					

3.3.1 A general time-to-event meta-analysis

In a Bayesian meta-analysis model for a time-to-event outcome, the quantity of interest is assumed to be a survival probability, i.e. the proportion of the population that have not had the ‘failure’ event [56]. Let r_{im} be the number of individuals in a study $i = 1, \dots, N$ that have not had the failure event at the end of time $m = 1, \dots, T$. Let n_{im} be the number of individuals in study i at risk at time $t - 1$. Then for each study:

$$r_{im} \sim \text{Bin}(n_{im}, p_{im})$$

$$p_{im} = f(\Phi)$$

$$\Phi \sim [.,.]$$

where p_{im} is the probability of those in study i not having a failure event by time m given they are at risk at time $m - 1$. This probability of failure is represented by a function $f(\Phi)$ defined by the chosen parametric survival model. The parameters in Φ are given prior distributions (Section 1.2.3).

3.3.2 Meta-analysis for modelling adherence to interventions

Bayesian time-to-event meta-analysis models, as defined in Section 3.3.1, were fitted to data consisting of the number and denominator of patients remaining adherent at one or more times for multiple studies. Adherence was assumed to be binary - individuals are either adherent or non-adherent to their intervention. Ceasing use of the intervention was considered to be the time-to-event outcome using the adherence definition provided in the papers closest to that in Shapiro and Shapiro (2010) - at least four hours usage for more than 70% of nights [195]. Those who stop using their intervention are assumed not to restart at a later point or switch to an alternative intervention.

Since there is some evidence from the literature review (Section 3.2) of a high initial level of non-adherence, Weibull models were used. Weibull models can be used to represent a change through time in the hazard of the event [117]. The probability density function of a Weibull survival model for a time-to-event, T is [117]:

$$f(t) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha)$$

with $\alpha > 0$ and $\lambda > 0$. The Weibull hazard function and survival function are, for $t \geq 0$:

$$h(t) = \lambda \alpha t^{\alpha-1}$$

$$S(t) = \exp(-\lambda t^\alpha)$$

If the shape parameter α is greater than one, the hazard increases with respect to time; if α equals 1, the hazard is constant with respect to time (and the model reduces to an exponential survival model); and if α is less than one the hazard decreases with respect to time [117].

For interpretation, the shape parameter can be represented as a function of the hazard ratio between two timepoints t_1 and t_0 ($t_1 > t_0$):

$$HR(t_0, t_1) = \left(\frac{t_1}{t_0}\right)^{\alpha-1}$$

Assuming $t_1 = 2t_0$, α can be expressed as a function of the hazard ratio for a doubling of time:

$$\alpha = 1 + \log_2(HR(t_1, t_0))$$

The scale parameter, λ is indicative of the mean survival time:

$$E(T) = \left(\frac{1}{\lambda}\right)^{\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)$$

$$\approx \frac{1}{\lambda}$$

when α is close to one. These interpretations of λ and α are useful when setting priors for the meta-analysis.

Recalling the notation presented in Section 3.3.1, with $N = 18$ and $T = 10$ then:

$$r_{im} \sim \text{Bin}(n_{im}, p_{im})$$

where p_{im} is the probability of an individual in study i remaining adherent at the end of time period m . p_{im} is related to the shape and scale parameters of the Weibull distribution for the time to non-adherence as follows:

$$p_{im} = \exp\left(-\lambda_i t_{im}^{\alpha_{j_i}}\right)$$

The scale parameter is assumed to be a random effect differing between studies i with:

$$\log(\lambda_i) \sim N(\mu_{j_i}, \sigma^2)$$

with a mean μ_{j_i} dependent on the treatment $j_i \in \{MAD, CPAP\}$ and a variance (σ^2) assumed common between treatments, given the limited data on MADs. The prior distributions on μ_{j_i} were dependent on the intervention:

$$\begin{aligned}\mu_{MAD} &\sim N(-5, 10^4) \\ \mu_{CPAP} &\sim N(-5, 10^4) \\ \sigma &\sim U(0, 10^4)\end{aligned}$$

implying a prior mean survival of $\approx e^5$ months ≈ 12 years, though with substantial prior uncertainty on the mean and the extent of heterogeneity in the scale parameters.

Four alternative assumptions for the shape parameter of the Weibull model (α_{j_i}) are considered here:

Model I: Shape parameter is the same for all studies and both interventions

$$\alpha_{j_i} = \alpha_{MAD} = \alpha_{CPAP} \sim U(0, 2)$$

implying a 95% prior credible interval for the hazard ratio due to a doubling of time of between 0.5 and 2.0.

Model II: Shape parameter is the same for all studies, but dependent on the intervention

$$\begin{aligned}\alpha_{MAD} &\sim U(0, 2) \\ \alpha_{CPAP} &\sim U(0, 2)\end{aligned}$$

indicating a 95% credible interval for the hazard ratio, due to a doubling of time, of between 0.5 and 2.0 for both interventions.

Model III: Shape parameter is a study specific random effect and dependent on intervention

$$\begin{aligned} \log(\alpha_i) &\sim N(v_{ji}, \omega^2) \\ v_{MAD}, v_{CPAP} &\sim U(-7, 7) \\ \omega &\sim U(0, 10) \end{aligned}$$

indicating the hazard ratio due to a doubling of time has a prior expected value of one with substantial prior uncertainty for both interventions.

Model IV: Shape parameter is a study specific random effects, dependent on intervention and conditional on the value of the scale parameter

This is equivalent to $(\log(\lambda_i), \log(\alpha_i))$ having a multivariate normal distribution.

$$\begin{pmatrix} \log(\lambda_i) \\ \log(\alpha_i) \end{pmatrix} \sim MVN_2 \left(\begin{pmatrix} \mu_{ji} \\ v_{ji} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\omega \\ \rho\sigma\omega & \omega^2 \end{pmatrix} \right)$$

This is a bivariate normal distribution, so it can be easily represented in terms of conditional distributions:

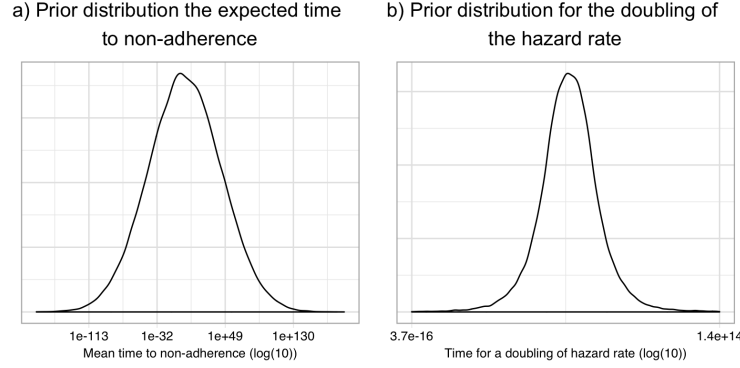
$$\begin{aligned} \log(\lambda_i) &\sim N(\mu_{ji}, \sigma^2) \\ \log(\alpha_i) | \log(\lambda_i) &\sim N\left(v_{ji} + \frac{\omega}{\sigma} \rho (\log(\lambda_i) - \mu_{ji}), (1 - \rho^2) \omega^2\right) \end{aligned}$$

where:

$$\begin{aligned} \mu_{MAD}, \mu_{CPAP} &\sim N(-5, 10^4) \\ v_{MAD}, v_{CPAP} &\sim U(-7, 7) \\ \sigma &\sim U(0, 10^4) \\ \omega &\sim U(0, 10) \\ \rho &\sim U(-1, 1) \end{aligned}$$

This has a similar interpretation to model III with the expected hazard ratio for a doubling of time being 1. The expected time until non-adherence is approximately 12 years, as weak prior information is assumed on the correlation (ρ) between α and λ . The prior distributions

Figure 3.3 The prior distributions for the mean time to non-adherence and the doubling of the hazard rate for Model IV¹.



¹Model IV assumes the shape parameter depends on the intervention and the study and is also correlated with the scale parameter.

of the mean time to non-adherence and for a doubling of the hazard rate, by simulating from the random effects distribution conditional on the simulated values for the hyperparameters, are shown in Figure 3.3

Bayesian inference was performed by MCMC, using Just Another Gibbs Sampler (JAGS) software in R [126, 169, 176]. Two chains were used with a burn-in period of 10,000 simulations with 100,000 simulations used in the analysis. Convergence was checked by inspection of the trace plots.

Pooling the estimates from the meta-analysis

Two alternative pooling methods for obtaining this estimated probability of adherence $\hat{p}_j(t)$ pooled over studies were calculated [81, 98, 203, 243]. The first method calculates the adherence probability from the model with Weibull shape and scale given by the means from the random effects distributions:

$$\hat{p}_j(t) = \exp(-\bar{\lambda}_j t^{\bar{\alpha}_j})$$

where $\bar{\lambda}_j = \exp(\mu_j)$ and $\bar{\alpha}_j = \exp(v_j)$ (or α_j dependent on the meta-analysis model specification). This assumes the target population for the decision is the same as the average study setting in the meta-analysis. This method may not be appropriate for this model, especially for adherence to CPAP where the population in the meta-analysis appears to have more

severe OSAHS than the CEA model population.

The second method uses the predictive distribution which is equivalent to estimating the probability of non-adherence for a new ‘hypothetical’ study. This ‘hypothetical’ study which represents the population of interest is assumed to be ‘sufficiently similar’ to the studies included in the meta-analysis with a probability of adherence over time of:

$$\hat{p}_j(t) = \exp(-\lambda_{NEW} t^{\alpha_{NEW}})$$

where $\lambda_{NEW}, \alpha_{NEW}$ are generated from the random effects distributions. In practice, the pooled estimate from a hypothetical study will be similar to using the random effects mean but will be less precise. This reflects uncertainty about where the hypothetical study may lie with respect to the random effects distribution and the uncertainty in the random effects variance.

Other options exist for summary estimates, some of which are outlined below [98, 203, 242]. An independent study specific estimate involves using only the data and estimate for a specific study with the decision population (i^{dec}) assuming the target population is the same as in i^{dec} [242]. This does not use the meta-analysis and is similar to the Sharples et al. (2014) CEA [198, 242]. This pooling method requires that an appropriate study is available.

Alternatively, a shrunken study-specific estimate assumes the population of interest is similar to all studies in the data (as in the predictive distribution) but one particular study has the population of interest. In this case, the study specific estimate for this study from the meta analysis can be used [243]:

$$\hat{p}_j(t) = \exp\left(-\hat{\lambda}_{i^{dec}} t^{\hat{\alpha}_{i^{dec}}}\right)$$

where $\hat{\lambda}_{i^{dec}}$ and $\hat{\alpha}_{i^{dec}}$ are the estimates for the shape and scale parameter for study i^{dec} . This method will be more precise than using study specific data as it ‘borrows strength’ from the other studies in the meta-analysis. Additionally, in the context of network meta-analysis (Section 4.3.1) consistency equations from the meta-analysis can be used to estimate the effects for comparisons not included in i^{dec} [242].

3.3.3 Goodness of fit measures

To assess which meta-analysis model fits the data the best, the deviance information criterion has been used [126, 204]. The deviance is defined for the modelling of data y , probability model $p(y|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, and prior distribution $p(\boldsymbol{\theta})$ as:

$$D(\boldsymbol{\theta}) = -2\log p(y|\boldsymbol{\theta}) \quad (3.1)$$

which in a Bayesian model has a posterior distribution as it is an explicit function of $\boldsymbol{\theta}$. An obvious candidate for a Bayesian measure of fit is to compare models in terms of the posterior mean deviance, $\bar{D} = E_{\boldsymbol{\theta}}(D)$. However, more complex models will fit the data better giving a smaller \bar{D} while their predictive ability may not improve. Therefore, it is preferable to have a measure of ‘model complexity’ to counteract \bar{D} [126, 204].

The effective number of parameters in a model, p_D , was derived by Spiegelhalter et al. (2002) as [204]:

$$\begin{aligned} p_D &= E_{\boldsymbol{\theta}|y} [-2\log p(y|\boldsymbol{\theta})] + 2\log p(y|\tilde{\boldsymbol{\theta}}(y)) \\ &= \bar{D} - D(\tilde{\boldsymbol{\theta}}) \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ is a ‘good’ plug in estimate of $\boldsymbol{\theta}$. Letting $\tilde{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|y]$ which is defined to be $\bar{\boldsymbol{\theta}}$, then:

$$p_D = \text{‘posterior mean deviance’} - \text{‘deviance of posterior means’}$$

where p_D is not invariant to reparameterisation [126, 204].

The measure of fit, \bar{D} , can be combined with the measure of model complexity to obtain the Deviance Information Criterion (DIC) which estimates the predictive ability of the model as [204]:

$$\begin{aligned} DIC &= \bar{D} + p_D \\ &= D(\bar{\boldsymbol{\theta}}) + 2p_D \end{aligned}$$

Plummer (2008) showed the DIC is an approximation of a penalised loss function using a cross-validation procedure [170]. The differences between models in terms of DIC are compared, not the absolute values of the DICs. There is no formal rule as to what is an

important difference in DIC [126, 204]. Lunn et al. (2013) show that when considering two models: model 1 and model 2, $\exp \left[\frac{DIC_1 - DIC_2}{2} \right]$ is similar to a likelihood ratio. A difference of ten gives a likelihood ratio of 148 and a difference of five a likelihood ratio of 12 [126, 170]. Therefore, a difference in terms of $DIC > 10$ could be said to definitely rule out the model with the higher DIC. Differences of DIC of between five and ten are substantial [204]. However, if the difference of $DIC < 3 - 5$, and the models make different inferences about the parameters, a model choice should not be made using DIC alone [126, 204]. Choosing a model purely on the basis of its DIC should be avoided. As stated by Spiegelhalter et al (2002), many other features of models should be considered when choosing a model other than the DIC alone [204]. These include the scientific plausibility of the model specifications and the robustness of the results.

3.3.4 Methods for the inclusion of the results from a meta-analysis on adherence to a cost-effectiveness analysis

The treatment effects in the case study CEA are based on RCTs which predominantly reported results on an Intention To Treat (ITT) basis [198]. In an ITT analysis patients are analysed according to their allocated treatment regardless of whether they received or completed the treatment. Since ITT effects from trials with up to one year of follow-up are used in the case study CEA, the treated cohort simulated by the CEA model is implicitly assumed to contain those who stop adhering in their first year of treatment. This assumes non-adherence in the RCTs used to estimate the treatment effect is typical of the studies included in the meta-analysis.

The probability that an adherent individual becomes non-adherent in the 12 months (one year) following t estimated from the meta-analysis, is used to estimate the additional proportion of people in the cohort who become non-adherent from year one onwards:

$$q_j(t) = \frac{\hat{p}_j(t+12)}{\hat{p}_j(t)}$$

The values for the adherence rates, $\hat{p}_j(t)$ are taken from the posterior distributions from the meta-analysis models using pooled study estimates. For each PSA sample and intervention a different MCMC sample from the joint posterior distribution for the shape and scale parameters are taken, i.e. using the best fitting model (Model IV) from Section 3.3.2. The corresponding values for $\hat{p}(t)$ are used in the CEA in place of the adherence point estimates previously used in the case study (Appendix B.3.5).

The impact of the adherence model being used for three alternative time horizons (five years, ten years and the lifetime of the model (65 years)) was explored due to the limited follow-up of the adherence data (Section 3.2). At the end of this period, those patients still adherent to the intervention were assumed to remain adherent for the remaining lifetime of the model (Appendix B.3.5).

3.3.5 The value of collecting further information on adherence to interventions

Value of information methods are used to assess whether it would be useful to collect further information on adherence to MAD and CPAP. There are number of different value of information measures which can be calculated (Section 1.1.3 and Chapter 2).

The EVPI is the expected cost of uncertainty in the CEA model and defined as the difference between the expected NMB under perfect and current information (Sections 1.1.3 and 2.3) [24].

The Expected Value of Perfect Partial Information for adherence to interventions

The EVPPI is the expected value of resolving uncertainty on a set of parameters and is useful to help prioritise future research on those parameters where more precise estimates will lead to greater health benefits [24]. The method and notation outlined by Strong et al. (2014) are used to calculate the EVPPI for further information on adherence (Section 2.3.2) [215].

Using the case study CEA, incorporating the results of the random-effects adherence meta-analysis for model IV (Section 3.3.2) the EVPPI for the following sets of parameters, relating to adherence which is of principal interest, have been considered :

- information on adherence to MAD (i.e. eliminating uncertainty in the means of the random effects for the shape and scale parameters of the Weibull model for adherence to MAD): $\phi = \{\mu_{MAD}, \nu_{MAD}\}$
- information on adherence to CPAP (i.e. eliminating uncertainty in the means of the random effects for the shape and scale parameters for adherence to CPAP): $\phi = \{\mu_{CPAP}, \nu_{CPAP}\}$
- information on adherence to both MAD and CPAP (i.e. eliminating uncertainty in the means of the random effect estimates for the shape and scale parameters for adherence to MAD and CPAP): $\phi = \{\mu_{MAD}, \nu_{MAD}, \mu_{CPAP}, \nu_{CPAP}\}$

In addition, the EVPPI for other parameters in the CEA, found to have a non-zero EVPPI in Section 2.6.2, have been calculated to explore the impact of the inclusion of the adherence model in the CEA. These parameters are:

- change in ESS due to treatment with MAD : $\phi = \{\Delta ESS_{MAD}\}$
- change in ESS due to treatment with CPAP : $\phi = \{\Delta ESS_{CPAP}\}$
- change in SBP due to treatment with MAD : $\phi = \{\Delta SBP_{MAD}\}$
- change in SBP due to treatment with CPAP: $\phi = \{\Delta SBP_{CPAP}\}$
- annual cost of CPAP: $\phi = \{Cost_{CPAP}\}$
- annual cost of MAD: $\phi = \{Cost_{MAD}\}$

The Expected Value of Sample Information for collecting information on adherence to interventions

The EVPPI provides an upper bound on the value of future research for a set of parameters, ϕ . However, a study collecting further information on ϕ will have a set sample size, say m (Sections 1.1.3 and 2.4). The EVSI provides information on the value of conducting a particular research design with a sample size of m (Sections 1.1.3 and 2.4) [24]. The expected benefits of research can be compared to the expected costs of undertaking the study. If the expected benefits exceed the expected costs of research, there is value in carrying out the study (Section 2.4.1). This work deals with the calculation of EVSI and does not consider costing the research designs.

This work uses the methodology of Strong et al. (2015) using PSA output and non-parametric regression [215, 216]. This method is outlined fully in Section 2.4.2, but briefly recall that it requires the information from study data, X , about the parameters of interest to be expressed as a low-dimensional sufficient statistic, $T(X)$. Two study designs have been considered using the results of the CEA incorporating the results of the adherence meta-analysis, for a range of study population sizes $m=10, 100, 500, 1,000, 5,000$:

Study design 1: A population of size m collecting information on adherence to an intervention at one timepoint

Information on adherence to an intervention may only be available at one timepoint (t_0). For example when information is collected retrospectively. In this case the data collected would

consist of the proportion of participants still adherent to the intervention at time t_0 defined to be $X = X(t_0)$.

The parameters in the PSA relating to adherence are the shape (μ_j) and scale (v_j) parameters for a Weibull distribution with intervention $j = \{MAD, CPAP\}$ using the meta-analysis model IV specification.

A new study would aim to reduce the uncertainty in these parameters by being added to the meta-analysis. The shape and scale parameters are not separately identifiable from a single observation of the proportion of the population adherent and so it is not possible to find a sufficient statistic for μ_j and v_j separately. Given m , the (as yet unknown) information from the study can be fully expressed by the number adherent at t_0 ($X_j(t_0)$), which arises from the posterior predictive distribution under the meta-analysis given by the binomial distribution integrated over the posterior distribution of $p_j(t_0)$:

$$X_j(t_0) \sim \text{Bin}(\text{size} = m, \text{prob} = p_j(t_0))$$

To estimate EVSI, a sample is drawn from this posterior predictive distribution, as follows: Firstly, the proportion of individuals adherent at time t_0 needs to be calculated. So, for each MCMC iteration, k :

$$\hat{p}_j^{(k)}(t_0) = \exp\left(-\lambda_j^{(k)} t_0^{\alpha_j^{(k)}}\right)$$

is calculated as the probability of adherence to intervention j at time t_0 , where $\lambda_j^{(k)} = \exp(\hat{\mu}_j^{(k)})$ and $\alpha_j^{(k)} = \exp(\hat{v}_j^{(k)})$ are draws from the posterior distributions of the predictive pooled estimates (under Model IV). Therefore, a simulation of the number of participants adherent at time t_0 can be taken from:

$$X_j^{(k)}(t_0) \sim \text{Bin}(\text{size} = m, \text{prob} = \hat{p}_j^{(k)}(t_0))$$

Following Strong et al. (2015) (Section 2.4.2) with $T(X) = X_j(t_0)$, a sample from the probability distribution of $INB(j, \theta)$ is regressed on the sample of $T(X)$ to estimate EVSI.

Study design 2: A repeated measures study for a population of size m collecting information on adherence to an intervention at T timepoints

Let the T timepoints at which we propose to collect information on adherence to intervention j be (t_1, \dots, t_T) where $t_s < t_{s+1}$, $s = 1, \dots, T-1$. Let $X_j(t_s)$ be the number of people in the population adherent to intervention j at time t_s . We assume:

$$X_j(t_s) \sim \text{Bin}(\text{size} = m, \text{prob} = p_j(t_s))$$

with

$$p_j(t_s) = \exp\left(-\lambda_j t_s^{\alpha_j}\right)$$

where $\lambda_j = \exp(\mu_j)$ and $\alpha_j = (\nu_j)$ are the ‘predictive new study’ pooled estimates under meta-analysis Model IV. The proposed new study will reduce uncertainty on the shape (α_j) and scale (λ_j) parameters of the Weibull distribution which is used to model adherence to intervention j . Let $\{x_0, x_1, \dots, x_T\}$ be the, as yet unknown, numbers of individuals adherent at times $s = 0, \dots, T$, $x_0 = m$ derived from the study.

To express the information provided by the study data about λ_j and α_j as a sufficient statistic, Maximum Likelihood Estimation (MLE) methods are used.

Recall that it is assumed the population level of adherence decreases monotonically over time and so $x_s \leq x_{(s-1)}$, then for $s = 1, \dots, T$:

$$x_{js} \sim \text{Binomial}\left(\text{size} = x_{(s-1)}, \text{prob} = \frac{\hat{p}_j(t_s)}{\hat{p}_j(t_{s-1})}\right)$$

Again, for each PSA sample k , $\hat{p}_j^{(k)}(t_s)$ for each $s = 1, \dots, T$ can be calculated as:

$$\hat{p}_j^{(k)}(t_s) = \exp\left(-\lambda_j^{(k)} t_s^{\alpha_j^{(k)}}\right)$$

The proposed study wants to find the number of the original m patients still adherent at t_s , given they were adherent at t_{s-1} . This will provide a dataset of the proportion of the original study population at risk at each timepoint. In addition, due to the monotonicity of adherence assumption, information on the adherence at t_s implicitly provides a lower bound for adherence at all t_{s-1} where $0 \leq (s-1) \leq s$.

For each PSA sample, $k = 1, \dots, K$ we can simulate data $\mathbf{x}^{(k)} = \{x_0^{(k)}, \dots, x_T^{(k)}\}$, the estimated number of individuals adherent to intervention j at each timepoint. From this data, we wish to find estimates $\hat{\lambda}_j^{(k)}$ and $\hat{\alpha}_j^{(k)}$ for each k using MLE methods.

The intervention j subscript is dropped for the remainder of this section for ease of readability. Using $x_s^{(k)}$ for $s = 1, \dots, T$, $k = 1, \dots, K$ and letting $q_s = \frac{\hat{p}(t_s)}{\hat{p}(t_{s-1})}$, the probability of remaining adherent at time t_s given adherent at time t_{s-1} , for a single timepoint. We want to express the data collected from the study as a sufficient statistic $T(X)$. This statistic can be defined as $\{\hat{\alpha}, \hat{\lambda}\}$ given the data X . Although the estimator is not available in a closed-form, we can numerically maximise the likelihood for any simulated dataset X . Hence, we can draw a sample from the posterior predictive distribution of $T(X)$ to employ in the Strong et al. (2015) method for estimating EVSI [216]. The likelihood is constructed as follows:

$$L(\alpha, \lambda | x_{s-1}, x_s) = \binom{x_{s-1}}{x_s} q_s^{x_s} (1 - q_s)^{x_{s-1} - x_s} \\ \propto q_s^{x_s} (1 - q_s)^{x_{s-1} - x_s}$$

So, for a set of timepoints t_0, \dots, t_T , the likelihood is:

$$L(\alpha, \lambda | x_0, x_1, \dots, x_T) \propto q_1^{x_1} (1 - q_1)^{x_0 - x_1} \dots q_T^{x_T} (1 - q_T)^{x_{T-1} - x_T} \\ = \prod_{s=1}^T q_s^{x_s} (1 - q_s)^{x_{s-1} - x_s}$$

and the log-likelihood is:

$$l(\alpha, \lambda | x_0, x_1, \dots, x_T) = \sum_{s=1}^T [x_s \log(q_s) + (x_{s-1} - x_s) \log(1 - q_s)]$$

Numerical methods are used to find $\hat{\lambda}_j$ and $\hat{\alpha}_j$. In this case, the BFGS update to Newton's method is used to numerically maximise the likelihood⁵. This MLE step is carried out for

⁵This is an extension of the Newton-Raphson method for finding a maximum by finding the values when the derivative equals zero (i.e. the roots of the equation). Briefly, one starts with an initial guess. The tangent of the function is found at that point. The x intercept of this tangent is found and taken to be the next guess (as it is a better approximation to the root). The process is iterated until reaching an appropriate degree of accuracy. The BFGS method is an approximation of this method which can be used when the second derivative of the function in question is too computationally expensive to compute at each iteration. This update provides an approximation to the second derivative (Hessian) which is an initial matrix which is updated at each iteration dependent on the starting value for each iteration.

each simulation giving two vectors $\hat{\alpha}$ and $\hat{\lambda}$, each of length K . $T(X) = \{\hat{\alpha}, \hat{\lambda}\}$ can be used in the non-parametric regression step in the calculation of EVSI.

3.4 Results

3.4.1 The time-to-event meta-analysis for adherence to MAD and CPAP as interventions for OSAHS

Summaries of the posterior distribution for the five-year probability of adherence for each model, each study, and fitted pooled values are presented in Figure 3.4. The pooled shape and scale estimates of the models obtained from the random effects mean are presented in Table 3.3 along with the deviance statistics.

Table 3.3 highlights the difference between the four model specifications. Model IV found a very strong negative correlation between the shape and scale parameter.⁶ In terms of the deviance statistics, Model IV appears to fit the data best. However, the difference in DIC compared to Model III is not substantial. Therefore, the decision to proceed on the basis of Model IV has taken into account the strong level of correlation found between the shape and scale which indicates it may be wise to include this correlation in the model.

As expected, the pooled estimates of the probability of adherence at five years from the predictive distribution are similar to the estimates using the random effects mean but more uncertain (Figure 3.4). This additional uncertainty is most evident in Models III and IV due to the between study heterogeneity assumed for the shape parameter.

Figure 3.5 shows the estimated adherence over time from the meta-analysis. The posterior medians in Panel (a) show little difference between models I, II and IV in estimating adherence to CPAP. However, there are some differences in the posterior medians between the model specifications when estimating adherence to MADs. Panel (b) shows the posterior distributions from Model IV, which fits the data best on the basis of DIC, using the predictive distribution. The difference between the MAD and CPAP models is small and in terms of uncertainty at the 75% level there is little difference. Although, the 95% credible intervals (not presented) are more uncertain around adherence to MAD due to the lack of data. Panel

⁶The random effects distribution characterises studies as either having a faster non-adherence rate through time indicating a shorter mean time to non-adherence or a slower rate of non-adherence implying a longer time to non-adherence.

(c) shows the difference between MAD and CPAP adherence. The adherence rates between the two interventions are similar. As time increases the median difference in adherence tends to zero, although the uncertainty around the difference in adherence increases. Finally, Panel (d) shows the posterior median and 95% credible intervals compared between the random effects mean and predictive distribution for Model IV. This indicates the median estimates are similar for both pooling methods. However, as expected and supported by the theory (Section 3.3.2) and Figure 3.4, there is more uncertainty around the estimates using the predictive distribution.

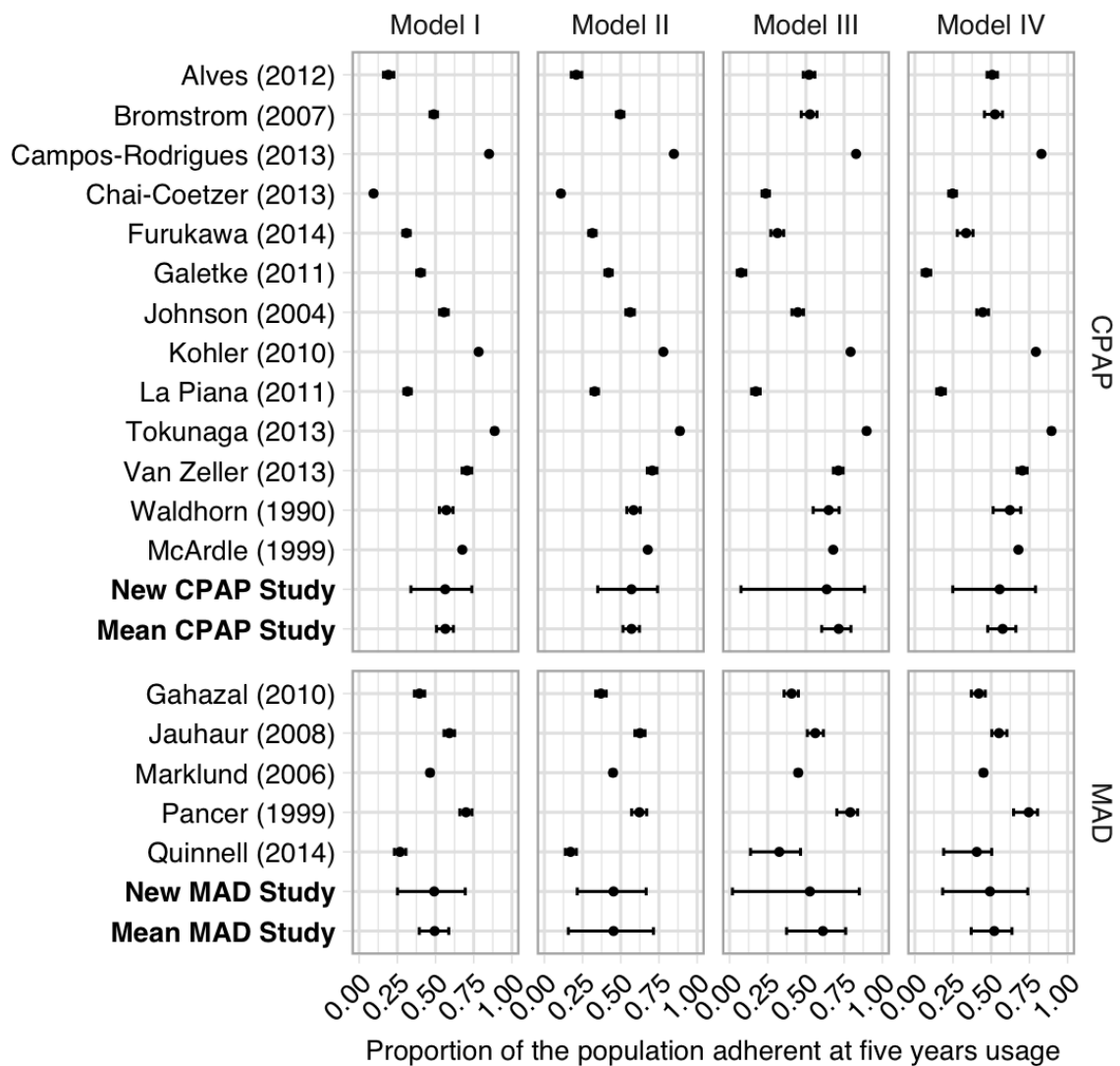
3.4.2 The cost-effectiveness analysis

The meta-analysis model with separate shape and scale parameters for each study, assuming correlation between the shape and scale (Model IV) was implemented into the case study CEA model (Section 1.3.3 and Appendix B). It was the model which fitted the data best, on the basis of DICs and the estimated correlation was strong, indicating this correlation should be modelled (Table 3.3). The adherence model was assumed to apply for five years or ten years after which time patients still adherent to the intervention are assumed to remain adherent for the remainder of the lifetime of the CEA model. A further analysis assumed the adherence model was assumed to apply for the lifetime of the model (65 years). - which may not be true as it requires the model to be extrapolated 55 years beyond the end of the data available. The base case CEA used point estimates for adherence to MAD and CPAP taken from Kohler et al. (2010) (Appendix B.3.5) [119].

There appears to be little difference in the results between the CEAs when the random effects mean study or a new predicted study was used to obtain the pooled estimates (Table 3.4). There is little difference in how cost-effective CPAP is comparing the time periods for the adherence model of five or ten years. This can be, in part, attributed to the small value for the shape parameter in the adherence model suggesting that as time increases, the rate of change in adherence rate decreases. This means the proportion of patients becoming non-adherent in each cycle reduces, leading to little change in the results of the CEA.

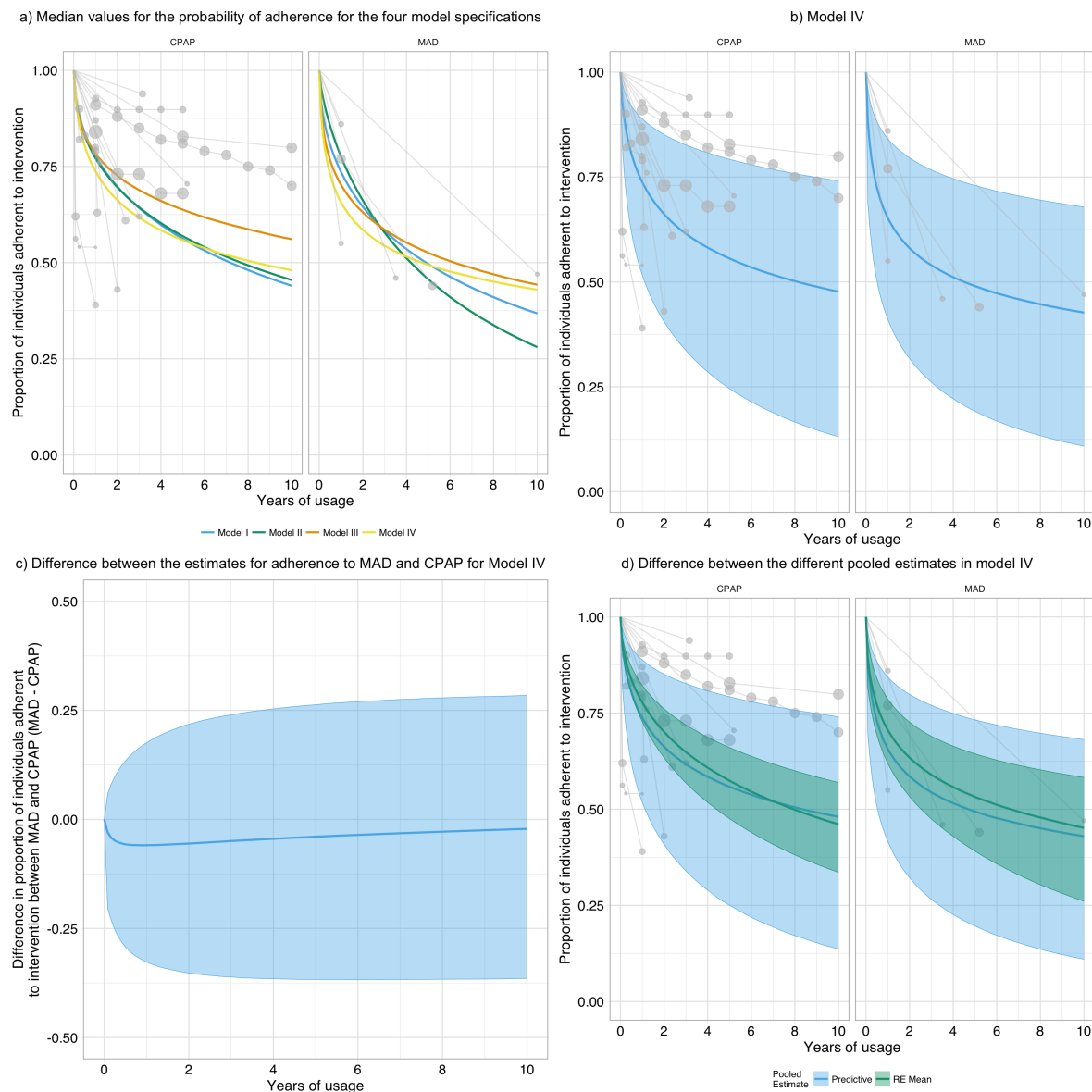
Compared to the base case, there is a very slight decrease in absolute terms in the NMB of MAD and CPAP when using the adherence model for ten years (the length of time data is available). This is due to the adherence model predicting lower adherence rates for both interventions than in Kohler et al. (2010) [119, 198]. There is more uncertainty around the optimal treatment decision when introducing the adherence model as illustrated by the CEAC and EVPI values (Table 3.4). This can, in part, be explained by the addition of further

Figure 3.4 Estimated five year posterior median and 95% credible interval adherence probabilities from four meta-analysis models¹ modelling adherence to MAD and CPAP as treatments for OSAHS along with the pooled estimates



¹all models assume that adherence to the interventions can be modelled with a Weibull distribution with a random effects scale parameters. Model I assumes the shape parameter is the same for all studies and both interventions. Model II assumes the shape parameter is dependent on the intervention but not the study. Model III assumes the shape parameter depends on both the intervention and the study. Model IV assumes the shape parameter depends on the intervention and the study and is also correlated with the scale parameter.

Figure 3.5 Observed adherence data¹ and pooled results with the 75% credible intervals where relevant from four Bayesian time-to-event meta-analyses of adherence to MADs and CPAP as treatments for OSAHS using 100,000 Monte Carlo simulations, using a predictive study pooled estimate unless specified



¹the points represent the actual data from the literature review. The size of the point is proportional to the size of the study. The lines between points from the same study are for ease of interpretation only; the proportion adherent between timepoints may not be linear in practice.

Table 3.3 Posterior median and 95% credible intervals for the Random Effect mean pooled estimates along with the deviance statistics from the Bayesian time-to-event meta-analysis models using 100,000 Monte Carlo simulations for adherence to MAD and CPAP as interventions for OSAHS

Parameter	Shape Parameter Distribution			
	Model IV Bivariate and different for all studies	Model III Different for all studies	Model II Different for each intervention	Model I Same for all studies
MAD				
Scale ($\bar{\lambda}_{MAD}$)	0.122 (0.016, 0.560)	0.104 (0.013, 0.525)	0.045 (0.018, 0.112)	0.087 (0.036, 0.205)
Shape ($\tilde{\alpha}_{MAD}$)	0.381 (0.034, 1.145)	0.349 (0.035, 1.297)	0.699 (0.594, 0.810)	0.515 (0.478, 0.549)
CPAP				
Scale ($\bar{\lambda}_{CPAP}$)	0.073 (0.029, 0.181)	0.073 (0.031, 0.167)	0.078 (0.045, 0.132)	0.070 (0.040, 0.120)
Shape ($\tilde{\alpha}_{CPAP}$)	0.488 (0.238, 0.814)	0.368 (0.161, 0.698)	0.483 (0.448, 0.521)	0.515 (0.478, 0.549)
Correlation ¹	-0.973 (-0.993, -0.863)			
Deviance Statistics²				
\bar{D}	653.4	654.4	877.4	892.0
pD	27.1	27.3	19.4	18.5
DIC	680.5	681.7	896.8	910.6

¹correlation between $\bar{\lambda}_j$ and $\tilde{\alpha}_j$ where $j \in \{\text{MAD, CPAP}\}$

²see Section 3.3.3 for further information on the deviance statistics.

uncertain parameters in the the CEA model. Additionally, under the adherence meta-analysis model, adherence to MAD is more uncertain relative to CPAP (Figure 3.5 Panel (c)). The treatment effect for MADs is smaller than CPAP (Appendix B). However, the impact of the increased uncertainty on adherence has the effect of the INB between the interventions to tend towards zero and the ICER to tend to the cost-effectiveness thresholds, increasing the uncertainty around the optimal treatment.

3.4.3 The value of information

The adherence model with separate shape and scale for each study was used, assuming correlation between the shape and the scale (Model IV) for ten years. The population-level EVPI estimates for all cost-effectiveness models and the base case are shown in Table 3.4. Using the adherence model increases the EVPI and hence the uncertainty in the optimal treatment comparison (as illustrated by the probabilities of being the most cost-effective intervention at a threshold of £20,000 per QALY gained in Table 3.4). All EVPPI and EVSI values have been calculated using the results of the CEA with uncertainty about adherence parameters expressed as the pooled estimates obtained from a new predicted study.

The EVPPI at a population-level for the adherence parameters and other parameters in the case study CEA are presented in Figure 3.6 These values have been calculated using the same population characteristics outlined in Section 2.2.2. This compares to a population-level EVPI of £95.8 million. All three population-level EVPPI estimates for parameters relating to adherence indicate potential value in conducting further research on adherence. The results indicate there is value in conducting further research around the other parameters in the model, though less value than for the adherence related parameters.

The EVSI at the population-level has been calculated for collecting information on adherence to MADs or CPAP at one or two timepoints (Figure 3.7). As the proposed follow-up time for patients' adherence increases so does the value of collecting the information. There appears to be value in collecting information for sample sizes as small as ten. As the sample size increases the EVSI increases, as expected, tending towards the population-level EVPPI values.

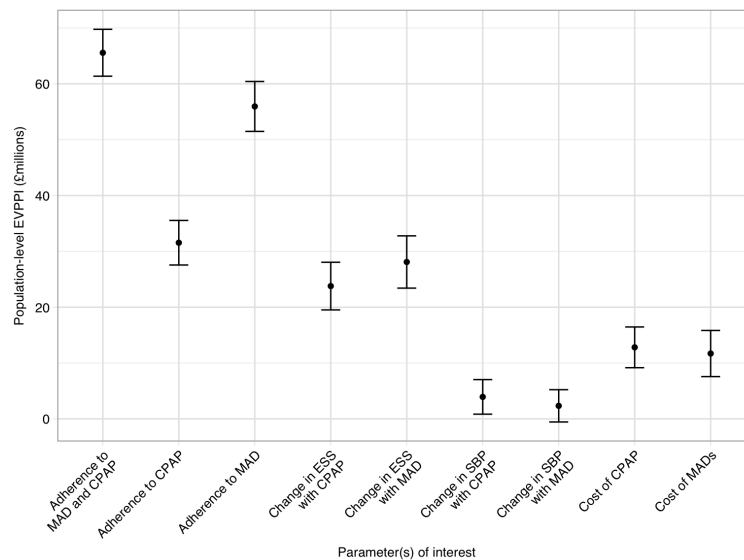
There is additional benefit in collecting information at points further into the future combined with information on adherence at one year. Collecting data for five and ten years usage of MADs for small sample sizes appears to be less valuable than collecting data at one and ten

Table 3.4 Results of the implementation of adherence Model IV¹ applied to the case study cost-effectiveness analysis using 100,000 PSA samples

	Base ²	Random Effects Mean			Predictive Distribution		
		Adherence Duration (years)			Adherence Duration (years)		
		5	10	65	5	10	65
LY							
CM	28.36	28.36	28.36	28.36	28.36	28.36	28.36
MAD	28.51	28.53	28.51	28.50	28.53	28.51	28.50
CPAP	28.61	28.59	28.56	28.54	28.59	28.56	28.54
QALY							
CM	14.35	14.35	14.35	14.35	14.35	14.35	14.35
MAD	14.65	14.62	14.59	14.58	14.62	14.60	14.58
CPAP	14.66	14.65	14.61	14.59	14.65	14.61	14.59
Cost (£)							
CM	6,112	6,121	6,120	6,108	6,121	6,116	6,112
MAD	8,331	7,243	7,187	7,411	7,250	7,204	7,450
CPAP	8,501	7,557	7,385	7,264	7,560	7,395	7,276
ICER							
MAD-CM	7,397	4,116	4,446	5,665	4,181	4,330	5,575
CPAP-MAD	7,239	10,467	9,900	-14,700	10,333	15,942	-17,400
NMB							
CM	280,888	280,879	280,880	280,892	280,884	280,879	280,888
MAD	284,669	285,157	284,613	284,159	285,150	284,796	284,150
CPAP	285,099	285,443	284,815	284,536	285,440	284,805	284,524
INMB							
MAD-CM	3,781	4,279	3,733	3,267	4,266	3,917	3,262
CPAP-MAD	430	286	202	377	290	9	374
CEA ³							
CM	0.0040	0.0002	0.0002	0.0004	0.0002	0.0001	0.0004
MAD	0.4366	0.4671	0.4691	0.4149	0.4814	0.4885	0.4326
CPAP	0.5594	0.5327	0.5307	0.5847	0.5184	0.5114	0.5670
EVPI ⁴ (£million)	63.8	84.5	92.1	76.7	86.5	95.8	80.0

¹The adherence model uses a different shape and scale parameter for each study, assuming correlation between the shape and scale parameters.²Model in the HTA replicated in R.³Cost Effectiveness Acceptability: the probability the intervention is the most cost-effective at a threshold of £20,000 per QALY gained.⁴Expected Value of Perfect Information: giving an upper bound for the potential value of future research, calculated at population-level using the methods in Section 2.2.2

Figure 3.6 Population-level¹ EVPPI quantities for the collection of data on adherence to MAD and CPAP as interventions for patients with OSAHS and other parameters using the case study cost-effectiveness analysis with adherence Model IV² for ten years and the new predicted study pooled values from the results of the Bayesian meta-analysis



¹Population-level values calculated using the parameters of the population-level EVPI presented in Section 2.2.2

²Model IV has a different shape and scale dependent on study and intervention with correlation between the shape and scale.

years usage. At larger sample sizes the values are comparable.

Comparing the EVSI for the interventions, the value of collecting information on adherence to MAD is greater than the corresponding value on adherence to CPAP apart from when collecting information on adherence at one year. The reduced EVSI for MADs relative to CPAP at one year may be due to a majority of the data on adherence to MADs having been collected at datapoints close to one years usage. In contrast, there is more information on longer usage with CPAP.

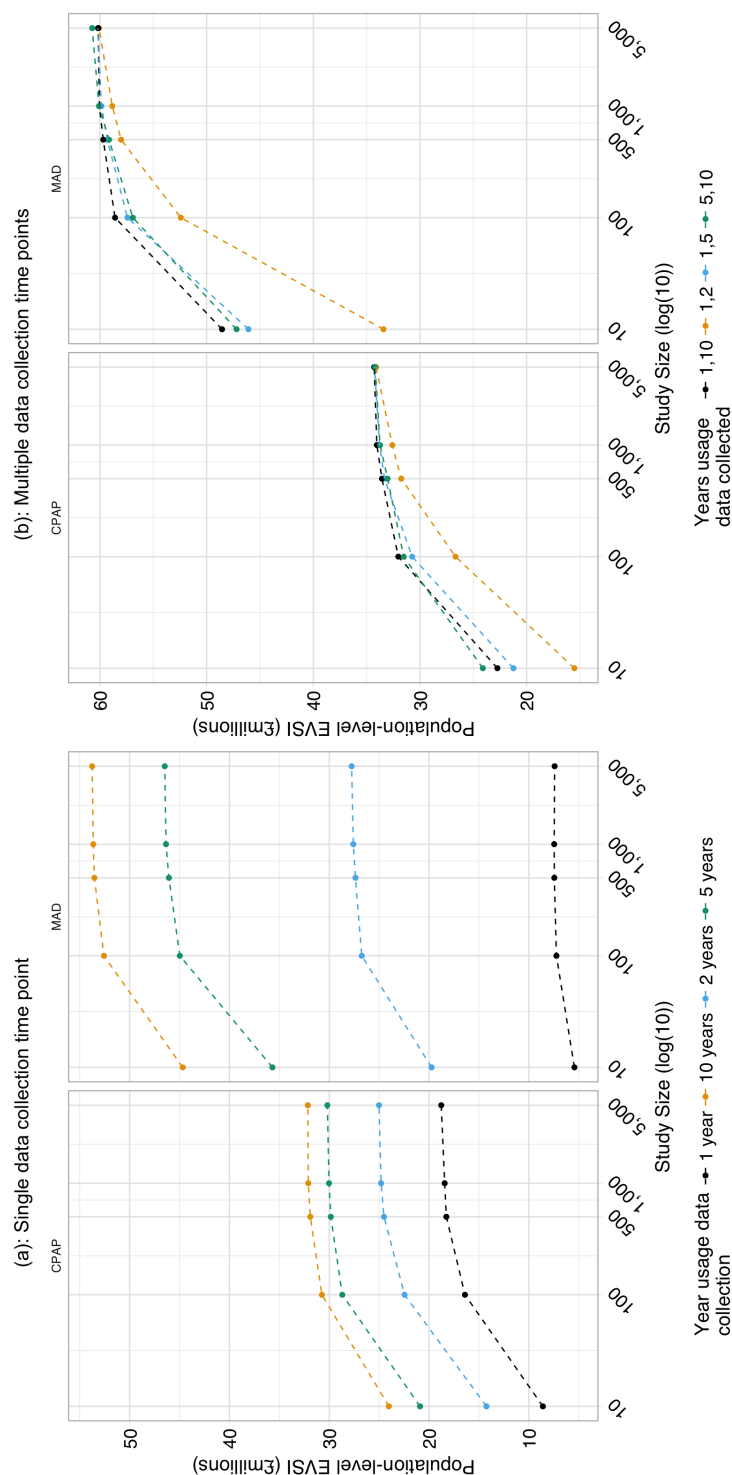
3.5 Discussion

Limited data is available on long-term adherence to MADs and CPAP as treatments for OSAHS with published studies using multiple definitions of adherence. The data suggests heterogeneity between studies in the populations explored is mainly due to the disease severity. The populations in studies examining adherence to CPAP were more severely affected by OSAHS than in those examining adherence to MADs, consistent with current NICE guidelines [149].

The chosen meta-analysis model for adherence (combining data on adherence to MAD and CPAP using a random effects Weibull model with different correlated shape and scale parameters for each study dependent on the intervention (Model IV)) fits the data best on the basis of deviance statistics (Table 3.3) although the difference in terms of DIC between the models with correlated and non-correlated shape and scale parameters (Models III and IV) was not substantial. The results from all the meta-analyses suggest adherence to CPAP is slightly higher, on average, than adherence to MADs (Figure 3.5 panel (c)), although there is additional uncertainty about adherence to MADs. Additionally, as the length of time using the interventions increases, the difference in the estimates of adherence between the interventions decreases.

Implementation of the meta-analysis using a random effects Weibull model with different correlated shape and scale parameters for each study (Model IV) into the case study CEA increased the uncertainty in the optimal treatment decision compared to the base case (Table 3.4), where adherence rates were taken as point estimates over a ten year period using Kohler et al. (2010) (Appendix B.3.5) [119]. Including modelled adherence to the case study CEA increases the EVPI, due to a difference in adherence rates between interventions, bringing

Figure 3.7 Population-level¹ EVSI estimated quantities for the collection of data on adherence to MAD and CPAP as interventions for patients with OSAHS at single timepoints (Panel a) and at two timepoints (Panel b) using the case study cost-effectiveness analysis with adherence model IV² for ten years and the new predicted study pooled values from the results of the Bayesian meta-analysis



¹Population-level values calculated using the parameters of the population-level EVPI presented in Section 2.2.2

²Model IV has a different shape and scale dependent on study and intervention with correlation between the shape and scale.

the effects of the interventions closer.

The EVPPI estimates indicated potential value in future studies looking at adherence to both interventions. However, as EVPPI assumes an infinite sample size the EVSI was calculated for the collection of further data of a specific sample size. Using the methodology in Strong et al. (2015), substantial value was found in collecting data on the number of people adherent to MADs and CPAP for sample sizes as small as 10. These gave population-level values in excess of £5 million [216]. There is additional value in collecting data on the proportion of the population adherent at two timepoints over collecting data at one timepoint for small samples sizes.

3.5.1 Data limitations

No formal assessment of the quality of the studies included in the literature review was made. Four studies were long-term follow-ups of RCTs [34, 82, 159, 174]. The remaining studies were prospective or retrospective long-term case studies analysing patients given the intervention at an earlier timepoint. The majority of studies used populations which were not (pre-)selected, indicating the data may be representative of usual practice. However, in many studies patients had to return questionnaires or attend an appointment to provide details on adherence which may suggest self-selection or social desirability bias among the population studied.

Of the 18 studies in the meta-analysis, seven examined adherence at one timepoint, and four considered adherence after over five years usage. No study reported information on adherence after ten years for either intervention. Therefore, an assumption had to be made about the time period the Weibull adherence model was valid over. In Table 3.4, the ICERs and the probabilities of being cost-effective at a threshold of £20,000 per QALY gained were similar when the adherence model was used for ten and 65 years (the lifetime of the case study CEA model). This is due to the shape parameter being small, reflecting a decreasing risk of becoming non-adherent leading to smaller year on year changes in adherence at later timepoints.

Three studies were excluded from the meta-analysis as they presented adherence as an average usage per night [36, 168, 217]. Converting the adherence values into binary measures is methodologically plausible by assuming usage has a normal distribution, $N(\mu, \sigma^2)$, using the mean and standard deviation of usage provided in the papers to calculate $P(\text{Usage} < X)$. X can be defined as the average number of hours usage a night, for studies where usage is defined as at least four hours a night for more than 70% of nights (assuming eight hours sleep

a night) [195]. This gives results which could be included in the meta-analysis models for the work by Chin et al. (2006) and Pieters et al. (1996) [36, 168]. However, Sucena et al. (2006) found the average usage per night increased over time, and this was not due to low users dropping out over time [217]. Therefore, using this transformation would give adherence increasing over time, violating the assumption made that patients can not become adherent having become non-adherent. On the other hand, if adherence is considered in terms of hours usage per night, this may increase over time as patients become more acclimatised to their intervention. However, an increase in usage per night is unlikely to change the definition of whether the patient is adherent under a specified binary measure of adherence.

The treatment effects in the case study CEA were taken from ITT estimates presented in RCTs with durations of less than a year. This means they were assumed to reflect the impact of non-adherence in the first year of usage. This may not be realistic. Many trials lasted significantly less than a year and the RCT populations may not reflect real life usage. However, the CEA used annual cycles, so assuming the treatment effect allows for non-adherence in the first year of use seems reasonable.

The meta-analyses assume adherence is binary. In reality, a treatment effect does not abruptly terminate once adherence falls below a certain level. In usual practice the treatment effect would diminish from full adherence to no usage. This is difficult to model and requires data not available in this case (information on the proportion of time people were asleep using/not using the intervention and the impact of different levels of adherence on the treatment effect). The case study CEA assumes non-adherent individuals are treated with CM which has no associated treatment effect (Appendix B.3.8). This may not be the case in practice. For example - if an individual is not adherent to MADs their clinician may start them on treatment with CPAP, as suggested by the NICE guidelines for patients with mild OSAHS [149]. A formal definition for adherence to interventions to patients with OSAHS was searched for and not found, therefore adherence is assumed to be more than four hours usage a night for 70% of nights - a common definition among a number of papers included in the meta-analysis [4, 77, 121, 195, 233]. A formal definition of adherence would be useful to guide further research.

The differences between studies in the literature review on adherence to interventions and the results of the meta-analysis may be driven by disease severity or patient preferences between interventions rather than the effect of the intervention. There is insufficient data to investigate this and those studies which looked at factors affecting adherence gave differing conclusions.

The differences in populations and the grouping of different types of MAD and CPAP used indicate results may not be well generalised [21]. In particular, information on how disease severity drives adherence and any difference in adherence to the different types of MADs and CPAP would be useful to explain the extent of heterogeneity.

3.5.2 Methodological issues

When implementing different definitions of the pooled estimates for the meta-analysis models, in the case study CEA little difference between the results was found. Different definitions of the pooled estimate represent different target populations for the treatment decision [97, 203, 243]. The decision population for the pooled estimates is that used in the case study - the TOMADO population (overweight, non-smoking, non-diabetic males with mild-moderate OSAHS who are moderately sleepy). The studies in the meta-analysis were similar to the target population in terms of age and BMI. However, the AHI was greater in the meta-analysis studies, especially in papers reporting on adherence to CPAP. However, AHI is not used in the case study CEA. We believe the study population is similar to the studies used in the meta-analysis, so the estimate from a predicted new study appears to be an appropriate pooled summary. Welton et al. (2015) outlined a number of other methods for pooling estimates which have not been explored, many dealing with heterogeneity between studies - assuming either independent study specific estimates, shrunken study estimates, or allowing for heterogeneity in the meta-analysis to be seen in the decision setting (Section 3.3.2) [243]. The application of these estimates to the work may be beneficial if more information was available to explain the causes of heterogeneity in the model.

The work in this chapter depends on the Weibull model being correctly specified. A three parameter model such as the Gompertz-Makeham model or a generalised gamma model could have been used [130, 206]. However, fitting such models requires sufficient data to identify more complex hazard variations. The Weibull model assumes the rate of becoming non-adherent is a monotonic function of time. A three parameter survival model would allow for non-monotonic functions. Since evidence from the literature and intuition suggest a high initial hazard reducing over time, therefore the exponential does not appear appropriate. Therefore, it was felt the Weibull model would be the most appropriate survival model for this data.

3.5.3 Future research priorities

In addition to collecting the information suggested of value by the EVPPI and EVSI calculations (Section 3.4.3) other data may help to improve the meta-analysis model. Data on adherence to CPAP for populations with mild-moderate OSAHS would help assess whether severity drives the difference in adherence between the two interventions. However, due to the current NICE guidance on treating OSAHS this data would be limited [149]. Data on adherence rates beyond ten years usage of MADs and CPAP would help determine the time period the Weibull model is applicable over and whether the Weibull model is appropriate. Due to the nature of the interventions patients may have strong preferences for a particular treatment. This may influence adherence, meaning information on patient preference between interventions and its relationship with adherence would be useful.

All data in the meta-analysis was published aggregated data. No IPD on adherence was available. IPD could assist in identifying factors such as AHI, ESS and BMI that may impact adherence rates. Additionally, IPD could help reduce uncertainty around meta-analysis estimates (Chapter 4).

The EVSI calculations do not take into account the costs associated with undertaking the proposed studies. To fully assess whether the proposed studies should be carried out the EVSI needs to be compared to the cost of undertaking the research (Section 2.4.1). Only if the value of the future work is less than its cost is it worth carrying out. Whilst we have not costed any of our proposed studies, we expect due to the population-levels estimated (Section 3.4.3) many of these studies would be worthwhile. The population-level EVSI estimates presented in Figure 3.7, are between £5-50 million depending on study size, the intervention considered, and the timepoints the data is collected at. These population-levels EVSIs appear to suggest these studies should be carried out. However, these values are dependent on the prevalence, incidence, and diagnosis rates of OSAHS over the time period of interest. As shown in Chapter 2, these are subject to uncertainty. Any change in the population can have a large impact on the value of information estimates.

The impact of the length of the proposed studies or the ease/difficulty of recruiting/collecting the data and participants have not been considered. A retrospective study, taking patients prescribed the interventions a number of years ago and seeing if they are still adherent is less informative than a prospective study which could collect data at multiple timepoints in the future. However, a retrospective study would provide the information quicker. It may be preferential to have data sooner rather than later as opposed to more valuable information.

This is not reflected in the EVSI calculations, and is a decision the researcher designing and implementing the new study needs to make.

The case study CEA uses a cohort who are all non-diabetic, non-smoking, overweight males with high blood pressure aged 50 with a baseline ESS value of 11.9 (Section 1.3). Running the model for different populations may yield different results in terms of the value of collecting information on adherence but would assume adherence was homogeneous. Additionally, the population in the case study CEA may impact the cost-effectiveness of the interventions regardless of whether adherence is modelled. Therefore, collecting data on treatment efficacy; adherence and other patient characteristics from a heterogeneous population would be useful, as discussed in Chapter 4.

The disease area and treatment options for the case study CEA could be considered a special situation. Both interventions for OSAHS are very different. For example, a CEA where both interventions are drugs may not have differential long-term adherence rates. Often many side effects due to pharmaceuticals leading to non-adherence occur soon after initiation of treatment and thus will be reflected in a treatment effect estimated from short term RCTs. Both treatments for OSAHS are currently available on the NHS although the current guidance on usage differs for the interventions [149]. This means there is a wealth of information over a number of years on both MAD and CPAP usage. This may not be the case for a treatment that is ‘new to the market’. In this case, available data would be more sparse than the data on MADs. This would make modelling adherence more difficult and the resulting estimated would have increased uncertainty. However, it would motivate estimating value of information quantities on adherence.

The methodology developed in this Chapter potentially has a wider scope than the application to adherence suggests. Further work exploring the applications of these methods to other binary time-to-event outcomes such as appropriate length of follow-up in a trial or for modelling the time-to-events in a CEA would be useful. These outcomes are potentially of more importance than adherence, and would expand the usage of the methodology.

3.6 Conclusion

There is limited data for long-term adherence to interventions for individuals with OSAHS. However, Bayesian time-to-event meta-analysis techniques can be used to model the impact of adherence over time and can provide inputs in a CEA which reflect all available data. This

meta-analysis represented the extent of uncertainty about variation in adherence over time and between people. Value of information methods found significant value, in excess of £5 million at the population-level for collecting more information on adherence. This value was particularly high for MADs and when the proposed study sample size was greater than ten with participants using the intervention for two years.

Chapter 4

Methods for stratifying the optimal treatment decision using non-binary measures of disease severity

Cost-effectiveness can differ between groups of the population for many reasons including the baseline characteristics of the patients. These differences in cost-effectiveness can be driven by the treatment effect and can lead to different optimal interventions for subgroups of patients. Exploration of the impact of continuous baseline characteristics on the treatment effect can help guide the decision of which intervention is best for which groups of the population. This chapter provides such an analysis using Bayesian meta-regression methods combining individual participant (IPD) and aggregate data (AD) from RCTs to determine the impact of non-binary baseline characteristics on the results of the case study CEA.

4.1 Introduction

Results from a CEA typically show whether an intervention is cost-effective ‘on average’ for the population studied [25, 84, 193]. This can lead to a suboptimal treatment recommendation for some patient groups [193]. Patients can be divided into many different subgroups using factors including individual baseline characteristics, such as the presence of prior disease or clinical measures, for example SBP [193]. Exploration of cost-effectiveness within subgroups can lead to a more efficient allocation of resources [193].

Sculpher (2008) outlined a taxonomy of possible sources of heterogeneity between study participants. He also addressed issues in estimating heterogeneity on model parameters and

how the associated uncertainty can be quantified [193]. The term ‘value of heterogeneity’ was introduced in Espinoza et al. (2014) as a measure to quantify the population-level health economic benefit of stratifying decisions by subgroups [67]. This concept is explored further in Chapter 5.

The NICE methods guidance (2013) provides some information on how to analyse data for stratifying decisions by patient subgroups in practice [150]. They believe exploring the cost-effectiveness of subgroup specific policies is important, ideally using pre-defined subgroups. The guidance indicates a preference for analysis using IPD and states the precision of the subgroup estimates should be reflected when exploring parameter uncertainty [150]. However, there appears to be no formal guidance on the methodology to be used to explore patient heterogeneity in CEAs. On the other hand, there are a number of NICE Technical support documents which deal with heterogeneity in treatment effects using meta-regression methods [55, 166]. These were designed to assist practitioners in methods for NICE Technology Assessments [55, 166].

Despite the obvious potential benefits of stratifying the treatment decision, and NICE’s recommendations, stratification is not regularly considered and/or implemented. Due to a likely lack of data, subdividing a population can lead to concerns with the quality of the data used to inform subgroup specific effects [44, 49, 87]. Often trials used in evidence synthesis have weak data on treatment-subgroup interactions. There are potentially additional costs to stratification which may be prohibitive and could eliminate the benefits of stratification. For example, stratification on a genomic marker would be costly to implement, so a large difference in the health benefits between interventions in the subgroups is required for this to be implemented. These additional costs are important, and methods to assess their impact are discussed further in Chapter 5.

A lack of sufficient data is one reason why stratification is not regularly implemented, although not considering stratification is the same as stating there is no value in stratification with complete certainty. However, there is often some potentially sparse AD available from published studies. The primary aim of this chapter is to apply Bayesian meta-regression methods using both AD and IPD to estimate the association of continuous (i.e. non-categorical) baseline measures with the treatment effect, and illustrate how a stratified CEA based on this analysis can be implemented. The results from the network meta-regression and the CEA can help decide whether the available data is sufficient to make decisions on stratification of optimal treatment at an individual patient-level. Chapter 5 expands upon this and looks at the

decision of stratification at a population-level. This complements the work by Espinoza et al. (2014), who explored the value of stratification on covariates [67].

IPD is the preferred data source for inclusion in a meta-regression, as recommended in the NICE guidance [150]. However, access to IPD is notoriously difficult [181]. A secondary aim of this chapter is to illustrate the benefit of even a small amount of IPD in improving the precision and accuracy of a Bayesian meta-regression for a treatment effect. This was also illustrated by Saramango et al. (2012) [190]. They developed a series of Bayesian network meta-analyses which allowed for both AD and IPD to explore the effectiveness of an intervention with binary covariates [190]. They found including even a small amount of IPD into the network meta-regression increased the accuracy of the treatment covariate interaction estimates, compared to using AD alone. Additionally, the inclusion of IPD reduced the inconsistency in the meta-regression.

This work is not the first to consider statistical methods to stratify the optimal treatment decision from a health economic perspective. Hoch et al. (2002) discuss how econometric techniques can be used to identify important subgroups [101]. They use a similar idea to Section 4.4, using patient covariates to adjust the NMBs to explore which factors impact the cost-effectiveness of an intervention. However, they used a CEA that used individual patient simulation, as opposed to the case study CEA, which was a cohort simulation. Phillippo et al. (2018) explored methods for adjusting results of network meta-analysis for treatment effects that vary between populations [167]. They explore two recent methods and acknowledge network meta-regression models that use both AD and IPD are attractive alternatives to their work.

This chapter proceeds as follows: Section 4.2 outlines the data available for exploring stratification on patients with OSAHS using continuous measures of disease severity [198]. Section 4.3 describes the methodology of evidence synthesis by meta-regression of AD and IPD to potentially guide stratification. It also describes the models fitted to the data collected in Section 4.2. Section 4.4 presents the methods used in incorporating the meta-regression results into the case study CEA. Additionally, an analytic representation of the expected relationship between the baseline characteristics and the CEA results is presented [198]. Section 4.5 reports the results of the meta-regression models introduced in Section 4.2 and applies these models to the case study CEA to assess whether cost-effectiveness differs for individuals in different strata [198]. Section 4.6 discusses the limitations of the models and data, along with future research priorities. Section 4.7 concludes the chapter and

links the findings of this work to Chapter 5, which assesses whether stratification should be recommended for a population accounting for the costs of stratification and the population distribution of the stratifiers.

4.2 A literature review to guide modelling the impact of a baseline disease severity measure on the treatment effects with an intervention for patients with OSAHS

The case study report and its associated meta-analysis paper (Sharples et al. (2014, 2015)) reported results from a number of classical meta-analyses of RCTs to, among other reasons, provide information for the input parameters in the case study CEA [198, 199]. Their work builds on previous literature reviews focussing on the effects of treatment with MADs on the AHI and ESS (Lim et al. (2009)); and of CPAP and its impact on AHI and/or ESS (McDaid et al. (2009)) [123, 134].

In this thesis, the review is updated to collect additional data from the RCTs on the association of baseline values of ESS with the difference in ESS between treatment arms (treatment effect). Sharples et al. (2014, 2015) classified each study as having a population with mild ($ESS < 9$); moderate ($9 \leq ESS < 15$) or severe ($ESS \geq 15$) daytime sleepiness at baseline [198, 199]. Subgroup meta-analyses were carried out for each categorisation of baseline ESS. No analysis looked at the explicit relationship between the continuous study-level baseline ESS and the study-level treatment effect. These studies reported their baseline ESS as an average for their population with a standard error. Having numerical, as opposed to categorical, values for the baseline ESS enables a clearer relationship of the association between the baseline ESS and the treatment effect to be estimated using meta-regression (Section 4.3.2). Thus, the papers were re-reviewed for information on the mean and standard error for baseline ESS and the treatment effect. Information on the baseline BMI for the study population, which will also be considered as a stratifier, was also extracted. The papers were also searched to see if they provided easily accessible IPD.

4.2.1 Search criteria and methodology

The updated data comes from studies included in the meta-analysis presented in Sharples et al. (2014) using the search terms in the Appendix of Sharples et al. (2014) [198]. Briefly, the search strategy was as follows: eligible studies included RCTs of adult OSAHS pa-

tients with at least one arm randomised to MADs or CPAP and a treatment duration of at least one week. Studies comparing two MADs, different types of CPAP, animal studies, non-randomised studies and those not published in English were excluded. The inclusion criteria included RCTs of adult patients with newly diagnosed or existing OSAHS of any severity. Studies where OSAHS was not the primary diagnosis, or where the sleep disordered breathing was predominantly associated with heart disease, stroke or dementia were excluded. The primary outcomes of the review were AHI and ESS. In this thesis the focus is on the ESS.

Since the publication of Sharples et al. (2014, 2015) one paper was retracted (Sharma et al. (2011)), so removed from the analysis [196, 198–200]. Additionally, updated values for Craig et al. (2012) and Gagnadoux et al. (2011) have been used [48, 78, 199]. There were some studies where the data was unable to be extracted in the form desired or inaccessible (18 papers); did not present the standard error for at least one of the baseline ESS or change in ESS (three papers) or the population was inappropriate (three papers) [17, 69, 183]. The papers included are detailed in Sharples et al. (2014, 2015), with exclusions as discussed above [198, 199].

4.2.2 Results

Complete data were found for 41 two armed and two three-armed RCTs, giving a total of 47 treatment comparisons. 31 studies compared treatment with CPAP to CM; eight studies compared treatment with a MAD to CM; and eight studies compared treatment with MAD and CPAP. This indicates, as with data on adherence to interventions (Chapter 3), the available data for treatment with MADs is sparse compared to treatment with CPAP.

Of those studies comparing CPAP with CM, five had a mildly sleepy population (measured by ESS), 25 had a moderately sleepy population, and one had a severely sleepy population at baseline. Those comparing treatment with MAD to CM all had a moderately sleepy population at baseline. One study comparing treatment with MAD and CPAP had a mildly sleepy population with the remaining seven studies having a moderately sleepy population at baseline.

The populations were aged between 44 and 59 for studies comparing CPAP and CM; between 45 and 59 for those comparing MAD and CM; and between 44 and 51 for those comparing MAD and CPAP. For reference, the case study CEA cohort was aged 50 years [198].

In terms of BMI, considered as a second stratifier later in the chapter, the average BMI in the populations studied comparing CPAP and CM is $27\text{-}37\text{kgm}^{-2}$; for those comparing MAD and CM is $27\text{-}43\text{kgm}^{-2}$; and for those comparing MAD and CPAP $26\text{-}33\text{kgm}^{-2}$. For reference the case study CEA cohort had a BMI of 31kgm^{-2} [198].

For a study providing IPD, indexed by i , assuming each participant p provided information on ESS at baseline (t_0) and at follow-up time (t_1) with intervention j , the effect of intervention j over time on the ESS, y_{ijp} , can be expressed as:

$$y_{ijp} = ESS_{ijp}(t_1) - ESS_{ijp}(t_0) \quad (4.1)$$

where $ESS_{ijp}(t)$ is the ESS of participant p in study i with intervention j at time t . In a study that provides AD the y_{ijp} are latent, and the information provided is:

$$y_{ij} = \frac{1}{P_{ij}} \sum_{p=1}^{P_{ij}} y_{ijp} \quad (4.2)$$

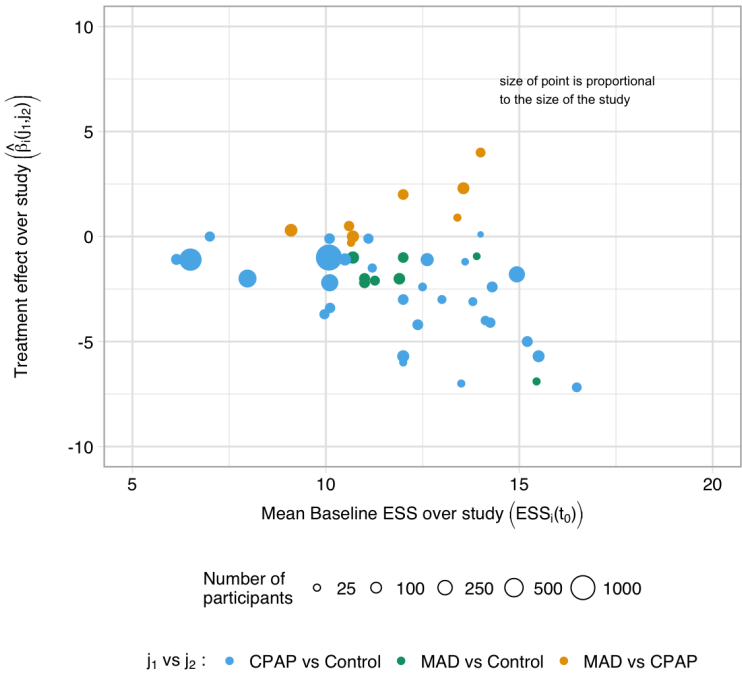
where P_{ij} is the number of individuals in study i treated with intervention j . The reported mean effect of treatment j_2 compared to j_1 in study i is:

$$\hat{\beta}_{i(j_1, j_2)} = y_{ij_2} - y_{ij_1} \quad (4.3)$$

Figure 4.1 shows the result of the literature review in terms of the mean study baseline ESS, averaged over all participants in all arms ($ESS_i(t_0)$) and the difference in mean ESS between treatment arms ($\hat{\beta}_{i(j_1, j_2)}$). This highlights the sparsity of the data relating to treatment with MADs. There appears to be some indication that as the baseline value of ESS increases, so does $\hat{\beta}_{i(MAD, CM)}$ and $\hat{\beta}_{i(CPAP, CM)}$. A conference abstract by Patel et al. (2017) found the minimum clinically important difference for the ESS was two, indicating more studies comparing CPAP and CM found a clinically important difference than studies comparing MAD and CM [160]. From Figure 4.1 there is little clear evidence on the impact of the relationship between the baseline ESS and the mean treatment effect over the study for studies comparing treatment with MAD and CPAP.

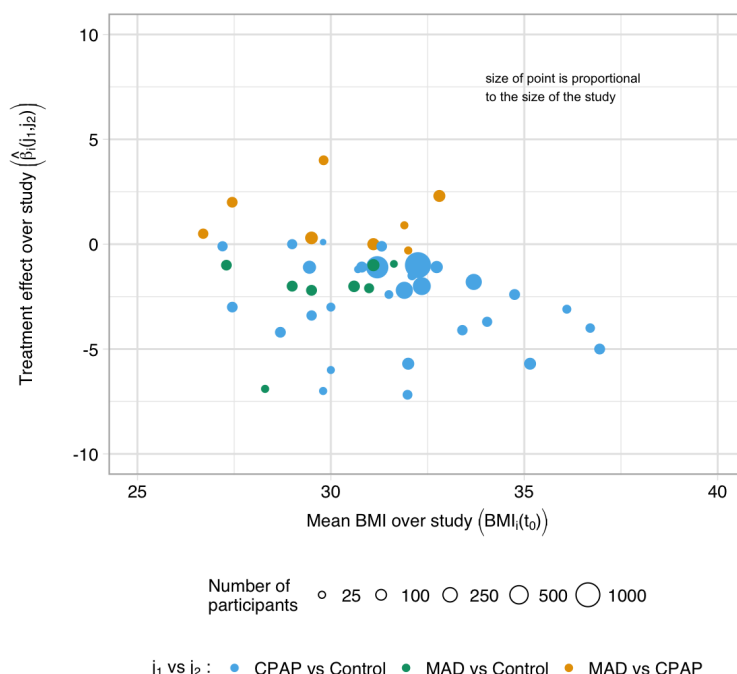
Figure 4.2 shows the relationship between the average BMI in the study ($BMI_i(t_0)$), at baseline, and the difference in ESS between treatment arms ($\hat{\beta}_{i(j_1, j_2)}$). This indicates some weak evidence that underlying BMI is associated with a stronger CPAP effect compared to CM. There is little evidence to support a relationship between underlying BMI and the effect

Figure 4.1 Observed data on baseline ESS values and the difference in mean ESS between treatment arms extracted from the studies in the literature review¹ [198, 199]



¹:A negative value indicates that intervention j_1 has a greater reduction in ESS than intervention j_2 .

Figure 4.2 Observed data on BMI values and the difference in mean ESS between treatment arms extracted from the studies in the literature review¹ [198, 199]



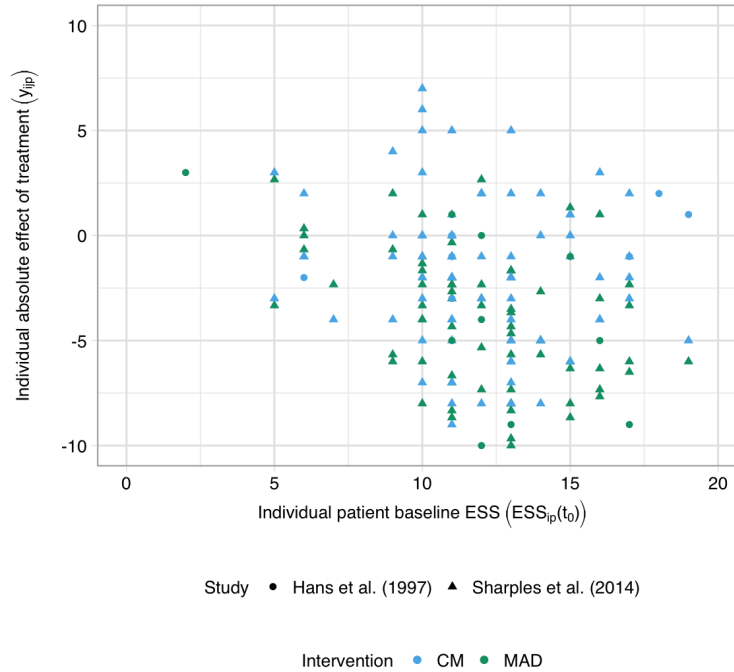
¹: A negative value indicates that intervention j_1 has a greater reduction in ESS than intervention j_2 .

of MADs compared to CM and the effect of CPAP compared to MAD.

Included above were AD from two studies where IPD was easily accessible. The TOMADO was the RCT the case study CEA is based upon [198]. The authors provided IPD for this study [198]. This compared treatment with a variety of MADs to no treatment. For this work, the values for the mid-range, semi-bespoke MAD have been used. In general, it is difficult to gain IPD from study authors. However, one study, Hans et al. (1997), a trial comparing treatments with MAD and CM with 21 participants, published their IPD [89]. This gave sufficient detail to be included.

Both Sharples et al. (2014) and Hans et al. (1999) were cross-over trials where participants receive multiple treatment. In both Hans et al. (1999) and Sharples et al. (2014) all patients receive treatment with MAD and CM, however, the order patients receive the interventions is randomised. This means for each patient there are two y_{ijp} , one for each $j \in \{MAD, CM\}$. Both studies implemented a washout period, a period of time between the periods of treatment where no treatment was given to allow for the effect of the first period of treatment to be

Figure 4.3 Observed data on the individual patient baseline ESS values and the individual patient difference in ESS with treatment over time extracted from the studies where IPD was easily accessible¹ [89, 198, 199]



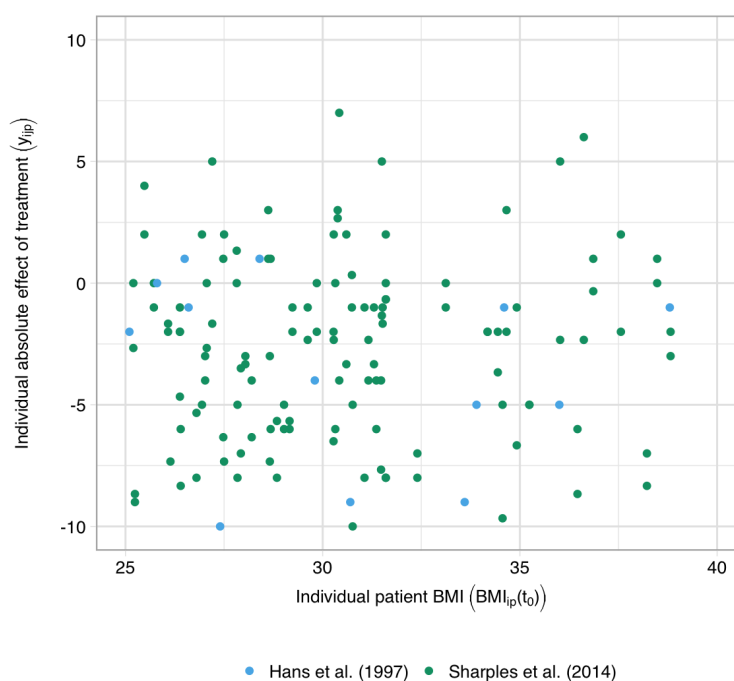
¹: A negative value indicates that intervention j_1 has a greater reduction in ESS than intervention j_2 .

eliminated. It has been assumed this washout period is sufficient to allow for the data to be treated as though it were from a parallel trial, with each patient providing two independent pieces of data.

The data from the studies providing IPD are presented in Figures 4.3 and 4.4. There is some evidence of a relationship between the baseline ESS and the change in ESS over time for each participant. This indicates an increase in baseline ESS may lead to a greater change in ESS over time when treated with MAD. The evidence of any corresponding relationship for BMI is weak. These relationships may also be due to those with a higher baseline ESS having a greater capacity to benefit.

It was decided to concentrate on absolute change in ESS over time as opposed to proportionate change to remain in line with the model and work in the case-study CEA [198].

Figure 4.4 Observed data on the individual patient baseline BMI values and the individual patient difference in ESS with treatment over time extracted from the studies where IPD was easily accessible¹ [89, 198, 199]



¹: A negative value indicates that intervention j_1 has a greater reduction in ESS than intervention j_2 .

4.3 Methods for estimating the impact of baseline measures on the treatment effect

4.3.1 Meta-regression theory

Meta-analysis methods, introduced in Section 1.2, combine multiple sources of data to estimate a single quantity, typically a treatment effect, for the population of interest [227]. Meta-regression is similar, but aims to link the effect size to one or more covariates [227]. Meta-regression recognises heterogeneity exists between studies and that the included covariates can help to explain some, but not all, of the heterogeneity [227]. This section introduces the meta-regression methods used in Section 4.3.2 by introducing methods for extending to network meta-regression, for allowing for trials with more than two arms, and for combining IPD and AD.

Motivating theoretical example

Assume n covariates of interest and I studies indexed by $i = 1, \dots, I$. Let \mathbf{Z} be a $I \times (n + 1)$ matrix of study-level covariates (including an intercept), and $\boldsymbol{\gamma}$ be a $(n + 1) \times 1$ vector of coefficients, with $\hat{\beta}_i$ the published treatment effect for study i , assumed to have been generated by the true study specific effects, μ_i , and a sampling error, σ_i , often taken as the reported standard error:

$$\hat{\beta}_i \sim N(\mu_i, \sigma_i^2)$$

A meta-regression model describes how the true study specific effects, μ_i , are distributed depending on \mathbf{Z} :

$$\begin{aligned}\mu_i &\sim N(\mathbf{Z}_i \boldsymbol{\gamma}, \sigma_\mu^2) \\ \boldsymbol{\gamma} &\sim [-, -] \\ \sigma_\mu^2 &\sim [-, -]\end{aligned}$$

σ_i^2 reflects the heterogeneity of the true study specific estimate. $\boldsymbol{\gamma}$ and σ_μ^2 the between study variance need to be estimated. $[-, -]$ represents the prior distribution for the pooled treatment effect, covariate effects and heterogeneity (Section 1.2.3).

Covariate values are not randomised between trials unlike treatments within trials, therefore meta-regression can suffer from confounding bias [227]. The aim of the work in this thesis is to choose the best treatment for a subgroup, not to explain the underlying cause of subgroup differences. Therefore, confounding bias, while important to acknowledge, is not thought to be a serious problem. In a similar way, the relationship of the individual patient outcome with trial-level or arm-level average covariates (the across trial relationship) may be an inaccurate estimate of the relationship of the individual patient-level outcome with patient-level covariates (within trial relationship). This is known as the ecological fallacy [59, 140]. The relationship of the outcome with covariates across and within trials is explored below.

Regression to the mean occurs in any analysis of the effect of baseline values on the same outcome that defines the treatment effect. In this case, the measurement error in the covariate also appears in the dependent variable which can potentially lead to bias in the meta-regression results [227]. In this thesis the ESS is measured by a questionnaire which has measurement error. This questionnaire is used in both the dependent variable and the covariate. This issue is not specific to the ESS and occurs in all patient reported outcome measures. It would be preferable for the patients' treatment effect to be related to some objective, low error, baseline value, such as the AHI - an objective measure of OSAHS severity [197]. Schilling et al. (2017) found using pre-post quality of life measures significantly impacted the results of the CEA compared to using matched controls and attributed this to regression to the mean [191]. As alternative data is not available it is important to acknowledge potential regression to the mean. However, we believe the effect of regression to the mean will be less than the minimally clinically important difference for the ESS [160]. In addition, shrinkage to the mean occurs in all hierarchical models. However, as we have a reasonable amount of data, this is unlikely to be considerable in this thesis [126].

Other issues with meta-regression, highlighted by Thompson and Higgins (2002), include the need for a sufficiently large amount of data [227]. Analysis can only use data from studies where complete information on the dependent variable and the covariate(s) of interest are available¹ [227]. This may lead to studies which could be synthesised in a meta-analysis not being included in a meta-regression. This means the resulting pooled estimates from a meta-regression may be subject to reporting bias [227]. Covariates of interest, to be included in the meta-regression, should be pre-specified with a plausible relationship with the dependent variable. Additionally, the number of covariates should remain manageable [227].

¹Theoretically, if some information was available on each study, missing data methods such as multiple imputation could be used. However, in practice this is unlikely to be helpful.

Mixed treatment comparisons and network meta-regression

The motivating example, introduced above, assumes two interventions are being compared. Now assume three interventions, $j \in \{0, 1, 2\}$, with $j = 0$ the comparator. Information on the pooled effect for each pair of treatment comparisons can be found using a network meta-regression. This can be used to find information for treatment comparisons where there is no available data [227].

The term ‘mixed treatment comparison’ is a subset of the term ‘network meta-analysis’ where there are closed loops in the network, created either by direct or indirect evidence [111]. In a network meta-regression, a transitivity assumption is required - the effect of the covariates, $\gamma_{(1,2)}$, on the pooled treatment effect comparing interventions 2 and 1 need to be related in the following way [110]:

$$\gamma_{(1,2)} = \gamma_{(0,2)} - \gamma_{(0,1)}$$

Taking a study i with an observed comparison between interventions $j = 1$ and $j = 2$, $\hat{\beta}_{i,(1,2)}$ the random effects model can be expanded to:

$$\begin{aligned}\hat{\beta}_{i,(1,2)} &\sim N(\mu_{i,(1,2)}, \sigma_i^2) \\ \mu_{i,(1,2)} &\sim N(\mathbf{Z}_i \gamma_{(1,2)}, \sigma_{(1,2)}^2) \\ &\sim N(\mathbf{Z}_i (\gamma_{(0,2)} - \gamma_{(0,1)}), \sigma_{(1,2)}^2)\end{aligned}$$

where, using the formula for the variance of the sum of two independent random variables:

$$\sigma_{(1,2)}^2 = \sigma_{(0,1)}^2 + \sigma_{(0,2)}^2 - 2\rho_{(1,2)}\sigma_{(0,1)}\sigma_{(0,2)} \quad (4.4)$$

with priors:

$$\begin{aligned}\gamma_{(0,2)}, \gamma_{(0,1)} &\sim [-, -] \\ \sigma_{(0,1)}^2, \sigma_{(0,2)}^2 &\sim [-, -]\end{aligned}$$

where $\rho_{(1,2)}$ is the correlation between $\mu_{(0,2)}$ and $\mu_{(0,1)}$. Assuming homogeneous variances between treatment comparisons, as the interventions are of a similar type then $\sigma_{(1,2)}^2 = \sigma_{(0,2)}^2 = \sigma_{(0,1)}^2 = \sigma^2$. By substitution of the homogeneous variances into Equation 4.4 the

correlation between the relative effect of $j = 1$ and $j = 2$ compared to $j = 0$ is $\rho_{(1,2)} = 0.5$ [56]. Lu and Ades (2004) present methods for heterogeneous variances [125].

Multi-arm trials

The literature review (Section 4.2) found two studies that compared three interventions. Methods to incorporate these while accounting for between arm dependencies within studies, are needed. Assuming exchangeability between studies (Section 1.2), a multi-arm trial, with $a > 2$ arms, produces a vector of $a - 1$ random effects $\boldsymbol{\mu}_i$. In this thesis there are three interventions, but this method can be extended to any dimension of a . Let the three interventions in study i be $j = 0, 1$ and 2 with study specific effects $\mu_{i,(0,1)}$ and $\mu_{i,(0,2)}$. Assuming the comparisons have a constant variance, σ^2 , the marginal distributions of the treatment effects are [97]:

$$\begin{aligned}\mu_{i,(0,1)} &\sim N(\mathbf{Z}_i \boldsymbol{\gamma}_{(0,1)}, \sigma^2) \\ \mu_{i,(0,2)} &\sim N(\mathbf{Z}_i \boldsymbol{\gamma}_{(0,2)}, \sigma^2)\end{aligned}$$

As the two study specific effects are relative to the same comparator, $j = 0$, correlation between them exists, so our interest is in the joint distribution of $\boldsymbol{\mu}_i = (\mu_{i,(0,1)}, \mu_{i,(0,2)})$ [56, 97, 125, 188]. Assuming transitivity, the marginal variance of $\mu_{i,(1,2)}$ is the same as the marginal variance of $\mu_{i,(0,1)}$ and $\mu_{i,(0,2)}$. Therefore, the covariance between $\mu_{i,(0,1)}$ and $\mu_{i,(0,2)}$ can be estimated using the marginal distributions above, the transitivity assumption, and the equation:

$$\begin{aligned}\text{var}(\mu_{i,(1,2)}) &= \text{var}(\mu_{i,(0,1)}) + \text{var}(\mu_{i,(0,2)}) - 2\text{cov}(\mu_{i,(0,1)}, \mu_{i,(0,2)}) \\ \sigma^2 &= \sigma^2 + \sigma^2 - 2\text{cov}(\mu_{i,(0,1)}, \mu_{i,(0,2)}) \\ \text{cov}(\mu_{i,(0,1)}, \mu_{i,(0,2)}) &= \frac{1}{2}\sigma^2\end{aligned}$$

The joint distribution of $\boldsymbol{\mu}_i$, for use in the network meta-regression when a study has three-arms is expressed as [56, 97, 125, 188]:

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i,(0,1)} \\ \mu_{i,(0,2)} \end{pmatrix} = MVN \left(\begin{pmatrix} \mathbf{Z}_i \boldsymbol{\gamma}_{(0,1)} \\ \mathbf{Z}_i \boldsymbol{\gamma}_{(0,2)} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{2} \\ \frac{\sigma^2}{2} & \sigma^2 \end{pmatrix} \right)$$

Incorporation of IPD into meta-regressions

In this thesis two studies which compared MAD and CM have easily accessible IPD [89, 198]. The aim of the meta-regression is to use all available data to assess whether the covariates impact on the change in ESS with treatment. Therefore, as IPD provides more information than AD, this data should be incorporated into the analysis.

A meta-regression using only IPD is seen as the ‘gold-standard’ and the ideal scenario for researchers. It enables information not available to AD only studies to be used [109, 182]. Within study relationships can be disentangled from between study effects [109, 182]. Further, IPD on patient-level covariates mean a network meta-regression can model the within study variation of effect modifiers. Additionally, confounding bias due to patient-level characteristics can be minimised and better subgroup estimates can be found through an IPD meta-regression [109].

However, it is unusual for a researcher to have access to IPD on all studies. If IPD were only available for a proportion of the studies, an IPD only meta-regression may be biased if the availability of IPD is related to the study results (reporting bias) [208]. Therefore, a meta-regression that uses all available data by combining both AD and IPD is seen as an effective and realistic compromise between a meta-regression of AD alone and one with full IPD. AD can estimate treatment effects on study-level factors for heterogeneous treatment effects, but IPD is needed to estimate the treatment effect on patient-level factors with homogeneous treatment effects [192]. Debray et al. (2016) explored in which situations IPD would be most beneficial. They concluded IPD would be most useful when included trials have substantial drop-out rate or the treatment effect appears to be dependent on patient-level covariates [52].

Riley et al. (2007) carried out a systematic review looking at meta-analyses that combined AD and IPD [181]. They found 33 articles which combined AD and IPD and 166 that did not, but described the results from IPD. The proportion of studies providing IPD ranged from 10% to 92% with a mean and median value of 64% and 71% across the set of 33 studies with no clear explanation for why some studies were able to collect more data than others. This represented 11% - 98% (mean and median of 67% and 68%) of total participants included in the meta-analyses.

Both Riley et al. (2007) and Debray et al. (2015) reviewed methods for combining AD and IPD in meta-analyses [51, 181]. Three distinct methods were found for continuous outcome measures, summarised below: a two-stage meta-regression model, partial reconstruction of

IPD, and multi-level modelling. Additionally, these papers identified some methods that can only be used for binary outcomes which have not been explored in this thesis as the focus is on non-binary outcomes.

Two-stage meta-regression method

Riley et al. (2007) found the two-stage meta-regression method, named by Simmonds et al. (2005), was the most commonly used method [181, 202]. In a two-stage method, for each study the IPD is reduced to AD. It is then combined with the existing AD using standard meta-analysis or meta-regression techniques. This method ignores patient-level information provided by IPD and so is best used when the predictors of interest for the pooled effect are at study-level [181]. Two-stage methods cannot be used if correlation exists between the pooled effect and the covariates of interest [55]. In addition, if the purpose of the meta-analysis is to look at effect modification by individual-level covariates then ecological bias may be introduced [51].

Dias et al. (2011) suggest the two-stage method could be used if the analyst does not have access to the IPD for some studies [55]. Data owners could calculate the summary values for their study, and provide these to the researcher who is carrying out the meta-regression. Further, Riley et al. (2007) state AD provided by the data-owners is likely to be of better quality than published AD [181]. However, as Dias et al. (2011) note, coordinating an analysis of this form could be cumbersome [55].

Reconstruction of IPD

This method can be used when the outcome is binary, ordinal, or survival based and so its use can be limited [51]. In this method summary information can be used to generate patient-level information [230]. This data can then be used in a IPD only meta-analysis. However, as patient-level covariate information cannot be extracted in this way, this method can only be used to combine IPD and AD in a meta-analysis without covariates.

Multi-level or hierarchical modelling

In this method, a joint likelihood with parameters shared between the AD and IPD can be built so all studies are used to estimate the overall treatment effect and the study-level covariates [51]. However, only the IPD gives strong information on the effects of patient-level covariates [51]. This method is also referred to as multi-level modelling, shared parameter

modelling or hierarchical related regression [51, 62, 181, 182]. This can be implemented in a number of ways. Riley et al. (2007) and Donegan et al (2013) estimate a single multi-level regression model using a dummy variable to indicate whether a study provides IPD or AD [62, 181]. Alternatively Jansen (2012), Thom et al. (2015) and Sutton et al. (2008) jointly estimated two related regression models for IPD and AD by Bayesian inference using MCMC methods [109, 221, 225].

Riley et al. (2008) use classical methods including restricted maximum likelihood (Section 1.2), whereas the work in this thesis takes a Bayesian approach. A number of studies outline a Bayesian approach for combining AD and IPD for binary outcomes and present applications of their methods [62, 109, 221]. Saramango et al. (2012) use this method to carry out a network meta-regression for a binary outcome, where the inclusion of even a small amount of IPD led to more accurate estimates for the treatment-covariate interactions [190].

Most work published to date has concentrated on binary outcomes which is not the focus of this thesis. However, as noted in Thom et al. (2015), these methods are easily extendible to continuous outcomes [225]. In their work, Thom et al. (2015) carried out a Bayesian network meta-analysis which combined AD and IPD using a mixture of study designs [225]. They found the inclusion of IPD had little effect on the results of their work. Ravva et al. (2014) used a Bayesian hierarchical model for a continuous outcome response using work by Jackson et al. (2006) as a basis by presenting a linear approximation for defining the AD model in terms of the IPD model [105, 178].

4.3.2 Network meta-regression models to estimate the impact of baseline measure of disease severity on the change in ESS with treatment for OSAHS

Using the data collected in Section 4.2 and methods introduced in Section 4.3.1 network meta-regression models for the association of two covariates with the absolute treatment effect, $\hat{\beta}_{i(1,2)}$, have been explored. The first model explores the impact of baseline ESS on $\hat{\beta}_{i(1,2)}$ with the second exploring the impact of baseline ESS and BMI on $\hat{\beta}_{i(1,2)}$.

In the current CEA model BMI is not an input. However, it is thought a higher BMI leads to an increased risk of developing OSAHS and an increased risk of CVEs [198]. Additionally, from a methodological point of view, exploring two continuous stratifiers is interesting: Firstly, stratification on two parameters is more challenging on the data, as more data are

required to produce estimates with an appropriate degree of accuracy; Secondly, due to the increased number of strata, two parameter stratification is more computationally demanding with many more PSA samples needed to make stratified optimal treatment decisions with sufficient confidence.

Throughout this work, it is assumed the relationship between the covariate(s) and the treatment effect is linear. This is a strong assumption discussed further in Section 4.6.

A method for a network meta-regression to guide stratification on one covariate

A meta-regression model combining IPD and AD for a single covariate is outlined below. Two studies have easily accessible IPD [89, 198]. These are both cross-over trials comparing MAD and CM. However, the method presented could easily be extended to trials using CPAP or with multiple arms. Each patient is assumed to provide two independent observations comparing ESS under the two interventions with the baseline ESS. This should be reasonable as both studies used a washout period of no intervention between treatment arms to allow for the effect from the first treatment period to wear off. A number of studies make a similar assumption. Elbourne et al. (2002) undertook a review of 184 meta-analyses which incorporated cross-over trials [65]. Of these 6% excluded cross-over trials completely², 11% considered cross-over trials separately, 52% considered data from the first period only, 30% considered the data as though it came from a parallel trial, and 1% considered the data from the cross-over trial as paired data [65]. Therefore, the assumption in this work appears reasonable and in line with other practitioners, especially considering the washout period which is a common mechanism in OSAHS studies.

Formally, for the trials providing IPD, recall from Equation 4.1 that y_{ijp} is the observed change in ESS over time for an individual p in study i treated with intervention j , i.e.:

$$y_{ijp} = ESS_{ijp}(t_1) - ESS_{ijp}(t_0) \quad (4.5)$$

where t_0 and t_1 are the baseline and follow-up time respectively and $ESS_{ijp}(t)$ is the ESS for patient p in study i treated with intervention j at time t . The available data is presented in Figure 4.3. Let μ_{ijp} be the underlying treatment effect for patient p in study i with intervention j then:

$$y_{ijp} \sim N(\mu_{ijp}, \Phi^2) \quad (4.6)$$

²These papers states this in their review methods. Only one study specified a reason for this exclusion - that for this particular treatment a cross-over design was inappropriate

where Φ^2 represents the variance of the individual participant observations. Each patient, p , in study i contributes two observations under different treatments j . Let $i = 1, \dots, n$ index the studies with IPD in the meta-regression. Additionally, let $j \in \{CM, MAD, CPAP\} = \{0, 1, 2\}$. Further, let z_{ip} be the baseline value of the covariate of interest for patient p in study i (i.e. $ESS_{ijp}(t_0)$); z_i be the baseline value of the covariate of interest for study i (i.e. the mean of z_{ip} over participants); r_i be the mean change in ESS over time with $j = 0$ in trial i ; ϕ_{ij} be the mean treatment effect comparing treatment $j = 1, 2$ and $j = 0$ when $z_{ip} = 0$; and α_i be the mean change in response under $j = 0$ for a one unit increase in z_{ip} . All values of baseline covariates are assumed, going forward, to be centred around their mean values to reduce the posterior correlation between the intercept and gradient terms.

Let γ_A^j denote the across trials association between a one unit change in study-level covariate values z_i and the outcome μ_{ijp} . Let γ_W denote the within trial association between a one unit change in individual patient covariate values, z_{ip} , and the outcome μ_{ijp} . The cases $\gamma_A^j = \gamma_W^j$ and $\gamma_A^j \neq \gamma_W^j$ are considered separately to assess whether ecological bias is a potential issue. When assuming $\gamma_A^j = \gamma_W^j$, the within and across trial relationships are thought to be identical: confounding does not affect the across trial treatment covariate interaction and there is no ecological bias in the across trial interaction (Section 4.3.1). If $\gamma_A^j \neq \gamma_W^j$, it is not possible to make inferences about individual patients using the study-level data and ecological bias is present. Under the two cases μ_{ijp} is defined by:

Case 1: $\gamma_A^j \neq \gamma_W^j$

$$\mu_{ijp} = r_i + \phi_{ij} + \alpha_i z_{ip} + \gamma_W^j z_{ip} + \gamma_A^j z_i \quad (4.7)$$

Case 2: $\gamma_A^j = \gamma_W^j$

$$\mu_{ijp} = r_i + \phi_{ij} + \alpha_i z_{ip} + \gamma_A^j z_{ip} \quad (4.8)$$

The model specified above can be aggregated over patients within a study to produce a meta-regression model for studies providing AD. Recall that:

$$y_{ij} = \frac{1}{P_{ij}} \sum_{p=1}^{P_{ij}} y_{ijp} \quad (4.9)$$

is the average treatment effect over time for the P_{ij} patients treated in study $i = 1, \dots, n_2$ with intervention j . The y_{ijp} are latent and unobserved in studies providing AD. Additionally, let:

$$z_i = \frac{1}{P_i} \sum_{p=1}^{P_i} z_{ip} \quad (4.10)$$

be the mean baseline value of the covariate in study i over all P_i patients in study i . The meta-regression model for studies with two and three arms are presented separately. For two armed trials, the published effect of treatment j compared to the comparator j_0 is:

$$\hat{\beta}_{ij} = y_{ij} - y_{i0} \quad (4.11)$$

whose distribution is implied by Equations 4.6 - 4.8:

$$\hat{\beta}_{ij} \sim N(\mu_{ij}, \hat{\sigma}_{ij}^2) \quad (4.12)$$

except that the variances $\hat{\sigma}_{ij}^2$ are taken as the observed standard error from the published studies - the validity of this assumption is discussed further in Section 4.6; μ_{ij} is the underlying true treatment effect for study i with intervention j , where:

$$\mu_{ij} = \phi_{ij} + \gamma_A^j z_i \quad (4.13)$$

where z_i is centred. For studies not providing IPD the additional terms in Equations 4.7 and 4.8 cancel out when the equations are aggregated over the participants. In particular, γ_W^j is not present (if $\gamma_A \neq \gamma_W$) as trials providing AD cannot provide information on within study effect modifiers.

For trials with three arms the bivariate normal distribution for $(\hat{\beta}_{i1}, \hat{\beta}_{i2})$ (Section 4.3.1) is:

$$\begin{pmatrix} \hat{\beta}_{i1} \\ \hat{\beta}_{i2} \end{pmatrix} = MVN \left(\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{2} \\ \frac{\sigma^2}{2} & \sigma^2 \end{pmatrix} \right)$$

where μ_{ij} and β_{ij} are defined as for two armed trials and σ^2 is the observed standard error from the studies, assuming homogeneous variances.

Random Effects distributions

The random effects distributions for the model are, for $j = \{1, 2\}$:

$$\begin{aligned} r_i &\sim N(\nu, \lambda) \\ \phi_{ij} &\sim N(\rho_j, \tau) \\ \alpha_i &\sim N(\theta, \eta) \end{aligned}$$

with ϕ_{i0} being 0.

For three-armed studies, there is correlation between ϕ_{i1} and ϕ_{i2} which needs to be accounted for (Section 4.3.1). $\phi_i = (\phi_{i1}, \phi_{i2})$ is modelled as a bivariate normal distribution with conditional distributions assuming a correlation of 0.5³(Section 4.3.1) [97, 125]:

$$\begin{aligned} \phi_{i0} &= 0 \\ \phi_{i1} &\sim N(\rho_1, \tau^2) \\ \phi_{i2} | \phi_{i1} &\sim N\left(\rho_2 + \frac{1}{2}(\phi_{i1} - \rho_1), \frac{3}{4}\tau^2\right) \end{aligned}$$

Priors

To fully specify the network meta-regression model priors need to be defined for the hyper-parameters and covariate effects (Section 1.2.3). Vague priors have been used with their structure dependent on prior knowledge.

γ_A^j measures how much the absolute treatment effect, μ_{ij} , changes with a one unit change of study-level baseline value of ESS, z_i . Therefore, the prior for γ_A^j should be informed by plausible values for μ_{ij} . The mean for the prior was set to be zero, reflecting uncertainty on whether the impact of the stratification variable (z_i) on the difference in ESS between treatment arms (μ_{ij}) is positive or negative. In addition, as there is a lack of data on the MAD and CM ($j = 1$ and $j = 0$) treatment comparison and for γ_W^j the same priors are set for both

³As explained in Section 4.3.1, we assumed homogenous variances, τ , due to the interventions being of a similar type. This is combined with the assumption of transitivity - that the marginal variance of ϕ_{i1} is the same as for ϕ_{i2} . The covariance between ϕ_{i1} and ϕ_{i2} can be shown to be $\frac{1}{2}\sigma^2$ and so giving the correlation of 0.5

treatment with MAD and CPAP and for γ_A^j and γ_W^j .

The ESS takes integer values from zero (not sleepy) to 24 (extremely sleepy). We wish to find a credible interval for the effect of the baseline ESS on μ_{ij} . The maximum plausible range for y_{ij} is $(-24, 24)$. Therefore the effect on μ_{ij} of a change from high to low covariate values could range from -48 to 48 .

From this, we can determine a plausible variance, assuming this is a 99.95% credible interval for a 24 unit change in z_i as:

$$\sigma = \frac{(48 - (-48))}{\Phi^{-1}(0.9995) - \Phi^{-1}(0.0005)}$$

γ_A^j and γ_W^j represent the effect of a one unit change in z_i so, the estimate for σ^2 needs to be divided by 24, giving $\sigma^2 = 0.78$. The full list of priors is:

$$\begin{aligned} \gamma_W^j, \gamma_A^j &\sim N(0, 0.78) & j = \{1, 2\} \\ \rho_j, \theta, v &\sim N(0, 10^3) & j = \{1, 2\} \\ \tau, \Phi, \lambda, \eta &\sim U(0, 2500) \end{aligned}$$

Models explored

The aim of the network meta-regression is to make best use of all available data. Therefore, separate analyses have been carried out to demonstrate the impact of the inclusion of IPD. In addition, as there is no prior knowledge of whether it can be assumed that within and across trial interactions are the same (i.e. if $\gamma_A^j = \gamma_W^j$) this has also been explored. The models presented are:

1. all trials with AD
2. trials with AD and IPD and assuming $\gamma_A^j \neq \gamma_W^j$
3. trials with AD and IPD and assuming $\gamma_A^j = \gamma_W^j$

The difference between models 1 and 2 and models 1 and 3 show the impact of including IPD. The difference due to assuming $\gamma_A^j = \gamma_W^j$ is shown by comparing models 2 and 3. Goodness of fit measures (Section 3.3.3) can only be compared between models 2 and 3 as model 1 uses a different dataset.

Bayesian inference was performed by MCMC using the JAGS software [126, 169]. Two chains were used with a burn-in period of 10,000 simulations. 100,000 simulations were carried out. Convergence was checked by inspection of the trace plots.

A method for a network meta-regression to guide stratification on two covariates

The methodology for the network meta-regression on two covariates of interest is similar to when considering one covariate. Let all parameters defined in the one covariate case maintain their definition. In addition define:

$$z_i = \begin{pmatrix} z_i^{(1)} \\ z_i^{(2)} \\ z_i^{(1)} z_i^{(2)} \end{pmatrix}^T \quad \gamma_A^j = \begin{pmatrix} \gamma_{A1}^j \\ \gamma_{A2}^j \\ \Gamma_A^j \end{pmatrix} \quad \gamma_W^j = \begin{pmatrix} \gamma_{W1}^j \\ \gamma_{W2}^j \\ \Gamma_W^j \end{pmatrix}$$

where the superscripts in z_i identify the (centred) first and second covariates of interest. Γ represents an interaction term quantifying how the treatment effect changes for a one unit change in the product of the two covariates. The model structure is then identical to the one for one covariate but with z_i , γ_A^j and γ_W^j replaced by their vectorised form. The random effects distributions are as in the case of stratification by one covariate.

Priors

In setting priors for γ_A^j , γ_W^j , Γ_A^j and Γ_W^j the same theory was used as with one covariate as no further information was available to guide in the choice of priors:

$$\gamma_{A1}^j, \gamma_{A2}^j, \gamma_{W1}^j, \gamma_{W2}^j, \Gamma_A^j, \Gamma_W^j \sim N(0, 0.78) \quad j = \{1, 2\}$$

As we have little information, the interactions have been given the same prior as the other parameters. However, it may be that there is increased uncertainty around the interaction terms. All other priors retain their definitions.

Models explored

The same models, software, and number of simulations have been explored as in the network meta-regression to guide stratification on one covariate. This is due to the similar model structure and the same data sources being used.

Pooling the estimates over the studies

The resulting change in ESS with treatment j for an individual in study i with the covariate values z under the model is:

$$\mu_{ij} = \phi_{ij} + \gamma_A^j(z - \bar{z}) \quad (4.14)$$

where \bar{z} are the centring values for z . A corresponding pooled estimate representing an average over all studies has been obtained by plugging in the means for the random effects distributions from the meta-regression model into Equation 4.14:

$$\bar{\mu}_j = \rho_j + \gamma_A^j(z - \bar{z})$$

This method assumes the target population for the decision is the same as the average setting from the studies in the network meta-regression. Section 3.3.2 presents alternative methods for pooling the estimates in further detail along with their advantages and disadvantages.

4.4 Application of the results of the meta-regression to the cost-effectiveness analysis

The aim of the network meta-regressions (Sections 4.3.2) was to investigate whether different patients should be prescribed different interventions on the basis of their baseline ESS or BMI in terms of cost-effectiveness under current information. Therefore, the results of the network meta-regression need to be incorporated into the case study CEA.

The posterior samples of ρ_j and γ_A^j have been extracted from the network meta-regression considering stratification on one covariate and used in the case study CEA through the resulting posterior mean samples of the effect of intervention j compared to comparator $j = 0$ have been estimated by:

$$\bar{\mu}_j = \rho_j + \gamma_A^j(z_1 - 12) \quad (4.15)$$

where z_1 is the baseline ESS and $\bar{z} = 12$. Similarly, for the two parameter stratification the posterior samples of ϕ_j , γ_{A1}^j , γ_{A2}^j , and Γ_A^j have been extracted and the resulting posterior samples of the effect of intervention j compared to comparator $j = 0$ have been estimated by:

$$\bar{\mu}_j = \rho_j + \gamma_{A1}^j(z_1 - 12) + \gamma_{A2}^j(z_2 - 32) + \Gamma_A^j(z_1 - 12)(z_2 - 32) \quad (4.16)$$

where z_2 is the BMI and $\bar{z}_2 = 32$. The $\bar{\mu}_j$ estimates are used in the PSA in place of the estimates used in the case study CEA (Appendix B). These are presented in Equations 4.17 and 4.18 and were derived using the results of the classical meta-analysis in Sharples et al. (2014) [198]:

$$\mu_{MAD} \sim N(-1.61, 0.34^2) \quad (4.17)$$

$$\mu_{CPAP} \sim N(-1.62, 0.34^2) \quad (4.18)$$

Stratified sampling is implemented with the CEA run a number of times for each of the strata under consideration - i.e for each baseline ESS (integer values from zero to 24) for potential stratification on one covariate and each combination of baseline ESS and BMI (integer value from 24-38 kgm^{-2}) for potential stratification on two covariates. These are used in Equations 4.15 and 4.16 as covariates, baseline ESS (z_1) and BMI (z_2), to estimate the impact of the treatment on the ESS, μ_j .

This is applied to the case study CEA model that uses the optimal adherence model from Chapter 3. That is, the adherence model with a bivariate distribution between the shape and scale parameters for a Weibull survival model for ten years using the predictive pooled adherence parameters.

4.4.1 Exploring the relationship between the baseline ESS and the change in ESS between treatment arms analytically

This section looks at the analytic relationship between the baseline ESS through the change in ESS with treatment and the CEA results to explain how cost-effectiveness is expected to vary between strata in this example. The terminology used here is defined in Appendix B. Note the results of the CEA do not directly depend on BMI so it is not possible to analytically derive the impact of a unit change in BMI on the CEA. The baseline ESS and change in ESS with treatment impacts on the case study CEA in two ways:

1. through the estimation of the baseline utility (U_{base}), for each intervention which impacts on the expected QALYs
2. through the risk of a RTA, which influences the expected survival, QALYs and costs for each intervention

The impact of baseline ESS on QALYs

In the case study CEA, the baseline utility, U_{base} , for a cycle spent with intervention j is estimated as:

$$U_{base} = \alpha + \beta \times (z + \mu_j) \quad (4.19)$$

where α and β are the coefficients of a regression model mapping ESS to utility (see Appendix B.3.4), z is the baseline ESS of the participant, and μ_j is the change in ESS with intervention j compared to CM ($j = 0$). Further adjustments are made to U_{base} for CHD and stroke events which are not impacted by the ESS. The utility after a RTA is not affected by ESS and does not use U_{base} . The model assumes there is no effect on survival or costs due to ESS.

Under the meta-regression model (Section 4.3.2) we assume, for potential stratification on one covariate, the effect of intervention j in terms of change in ESS is:

$$\mu_j = \rho_j + \gamma_A^j(z - 12) \quad (4.20)$$

The INB (Section 1.1.1) is expressed as:

$$INB(\theta) = \lambda \Delta_E - \Delta_C$$

It can be shown the relationship between the INB and the baseline ESS (z) is linear. Substituting 4.20 into 4.19 gives:

$$\begin{aligned} U_{base} &= \alpha + \beta(z + \phi_j + \gamma_A^j(z - 12)) \\ &= \alpha + \beta(\phi_j - 12\gamma_A^j) + \beta z(1 + \gamma_A^j) \end{aligned}$$

This implies a one unit increase in the baseline ESS (z) leads to a $\beta(1 + \gamma_A^j)$ increase in U_{base} .

Expected QALYs for each j are calculated using the equations outlined in Appendix B.3.4. A one unit increase in ESS adjusts the NMB by an additive factor of:

$$\lambda \beta (1 + \gamma_A^j) \sum_{t=0}^{65} (1 + i)^{-t} \quad (4.21)$$

where i is the discount rate applied to the outcomes of the CEA, which NICE take to be 3.5%p.a. [150]. Thus, the NMB is linear in terms of baseline ESS. As INB is defined as the

difference in NMB between two interventions, the INB is also linear in the baseline ESS and change in ESS. Therefore, there exists a relationship between the baseline ESS and the incremental quality adjusted survival. The form of this relationship depends entirely on the form of the network meta-regression relating the treatment effect on ESS (μ_j) to baseline ESS (z), i.e. linear in this case.

The impact of baseline ESS on the risk of a RTA

The ESS is also used to estimate the probability of transition to the RTA event state. There is a lack of empirical evidence on the impact of MADs on RTAs, so this impact is assumed to be based on the impact of MADs on ESS [198]. The odds ratio for a RTA using MADs ($j = 1$) compared to treatment with CM ($j = 0$) is assumed to be:

$$OR_{MAD} = OR_{CPAP} \times \frac{\mu_1}{\mu_2}$$

Substituting Equation 4.20 into the above gives the ratio of the change in ESS due to treatment with CPAP ($j = 2$) compared to MADs ($j = 1$) as:

$$\frac{\mu_2}{\mu_1} = \frac{\rho_2 + \gamma_A^2(z - 12)}{\rho_1 + \gamma_A^1(z - 12)}$$

A one-unit change in baseline ESS (z) has a very small impact on this value. Therefore, it appears a change in z has a negligible impact on the risk of RTA.

Considering both areas where the ESS impacts on the case study CEA the relationship due to baseline utility dominates. Therefore, the relationship between the NMB and baseline ESS is expected to strongly depend on the form of the network meta-regression, which in this work has been assumed to be linear. The validity of this assumption is discussed in Section 4.6.

4.5 Results of the network meta-regression and potential stratification of the optimal treatment decision

4.5.1 Potential stratification of the optimal intervention on one covariate: Baseline ESS

Bayesian model-based network meta-regression results

Summaries of the posterior distributions for the meta-regression parameters for each of the three network meta-regression models are shown along with the deviance statistics in Table 4.1.

When only AD is used in the network meta-regression there is no clear relationship between baseline ESS and the treatment effect with MADs. There is much uncertainty, due to a lack of data, when comparing treatment with MAD and CPAP. There is evidence of a much stronger negative relationship between baseline ESS and the effect of treatment with CPAP: each unit increase of baseline ESS is associated with a 0.31 fall in the ESS. The inclusion of IPD has little impact on the posterior distributions when the treatment is CPAP. This is as expected, all available IPD came from studies comparing treatment with MAD and CPAP. However, the estimates of γ_A^{MAD} (and γ_W^{MAD} when $\gamma_A^j \neq \gamma_W^j$) indicate a strong negative relationship between the baseline ESS and treatment effect. This highlights the impact and benefit gained from including even a limited amount of IPD.

The case study CEA assumed the cohort had a baseline ESS of 11.9 and the change in ESS with treatment had a normal distribution with mean -1.61 and standard deviation 0.34 for treatment with MAD and a mean -1.62 and standard deviation 0.34 for treatment with CPAP (Appendix B.3.8) [198]. Substituting a baseline ESS of 11.9 into the results (Table 4.1) gives an estimate for the change in ESS when treated with MAD of -1.85 and when treated with CPAP of -2.39. This suggests the effect of treatment on ESS, as estimated by the meta-regression, is greater than in the case study CEA. While both the network meta-regression presented in this thesis and the Sharples et al. (2014) paper use the same data, on the whole, the inclusion of IPD, the correlation between arms in three armed studies, and including the trials which compared treatment with MAD and CPAP have all contributed to the increase in treatment effect [198].

To explore the impact of treating three-armed trials as a single trial with correlated effects between arms, as opposed to two independent two-armed trials, a brief additional analysis was run (not presented). This found adjusting for the correlation between arms helped reduce uncertainty by strengthening information on comparisons we do not know as much about. Most importantly, adjusting for between arm correlation ensured the data was represented appropriately.

To decide which of the network meta-regression models to apply to the case study CEA, a model using both AD and IPD is preferable as all available data should be used. This limits the choice to models 2 and 3. Considering the DIC (Section 3.3.3), model 3 has a marginally but not substantially lower DIC value (Table 4.1). The lack of a substantial difference in DIC between models indicates the decision of the best model should not be based on DIC alone [126, 204]. When $\gamma_A^j \neq \gamma_W^j$ the credible interval for γ_W^j lies inside that for γ_A^j and the proportion of available IPD is limited. Assuming $\gamma_A^j = \gamma_W^j$ makes more structural assumptions on the meta-regression model which we believe are plausible and so leads to fewer parameters needing to be estimated with greater precision. Therefore, the impact of the change in ESS with treatment on the result of the case study CEA and hence whether a stratified treatment regime could be implemented can be explored with greater accuracy. Additionally, we are unable to think of any confounding factors or possible ecological bias that may cause the between and within study effects to differ. Therefore, the posterior estimates from the model where $\gamma_A^j = \gamma_W^j$ (Model 3) are used for implementation into the CEA.

The impact of including the results of the network meta-regression on the case study cost-effectiveness analysis

The posterior estimates from model 3, using all available AD and IPD and assuming $\gamma_A^j = \gamma_W^j$ (Table 4.1) have been used in the CEA from Chapter 3 - the case study CEA updated with the results of the meta-analysis on adherence to interventions [198].

Figure 4.5 presents the posterior distributions of INB (NMB(j=CPAP, θ) - NMB(j=MAD, θ)) for each possible baseline ESS (integer values between zero and 24). It is expected the relationship between the baseline ESS and the INB is approximately linear (Section 4.4).

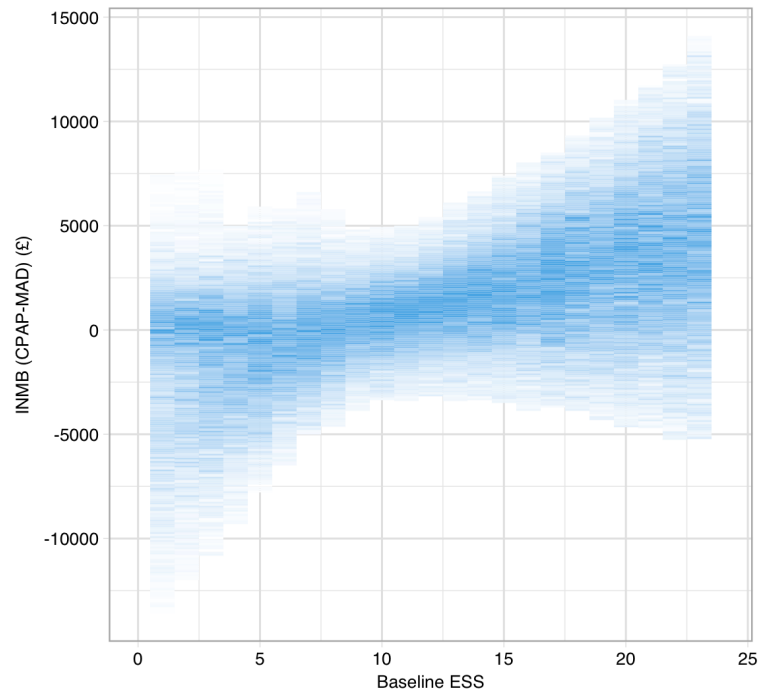
The optimal treatment changes when $E_{\theta}[INB(\theta)] = 0$. This occurs when the baseline ESS is seven (the ESS can only take integer values). This indicates that on the basis of NMB those patients with a baseline ESS of less than seven should be recommended treatment with MAD and those with a baseline ESS greater than seven should be recommended treatment with

Table 4.1 The posterior median, 95% credible intervals and deviance statistics for the coefficients of the network meta-regression models on the impact of baseline ESS on the change in ESS using aggregate^(a) and individual participant^(b) data with 100,000 Monte Carlo simulations and pooled effects as the means from the posterior distributions

	Model 1 AD	Model 2 AD + IPD $\gamma_A \neq \gamma_W$	Model 3 AD + IPD $\gamma_A = \gamma_W$
MAD^(c,d)			
ρ_{MAD}	-1.67 (-2.46, -0.02)	-1.77 (-2.55, -1.05)	-1.87 (-2.59, -1.18)
γ_A^{MAD}	0.03 (-0.49, 0.54)	-0.01 (-0.54, 0.49)	-0.17 (-0.43, 0.10)
γ_W^{MAD}		-0.23 (-0.54, 0.09)	
CPAP^(c,d)			
ρ_{CPAP}	-2.40 (-2.94, -1.92)	-2.42 (-2.97, -1.93)	-2.42 (-2.98, -1.18)
γ_A^{CPAP}	-0.31 (-0.52, -0.13)	-0.31 (-0.53, -0.14)	-0.32 (-0.54, -0.15)
Deviance Statistics^(e)			
\bar{D}	141.6	1,061.0	1,060.0
pD	17.2	23.4	22.9
DIC	158.8	1,084.0	1,083.0

(a)AD; (b)IPD; (c) ϕ_j = The pooled treatment effect for the mean baseline ESS; (d) γ^j = The effect of a one unit increase in individual (W) or average (A) baseline ESS on the treatment effect with intervention j ; (e)see Chapter 3 for further information on these quantities.

Figure 4.5 Results of the case study cost-effectiveness analysis incorporating the results of the network meta-regression on the impact of baseline ESS on the change in ESS with treatment^(a) presented in terms of the INB between MAD and CPAP for all values of baseline ESS and a cost-effectiveness threshold value of £20,000 per QALY gained

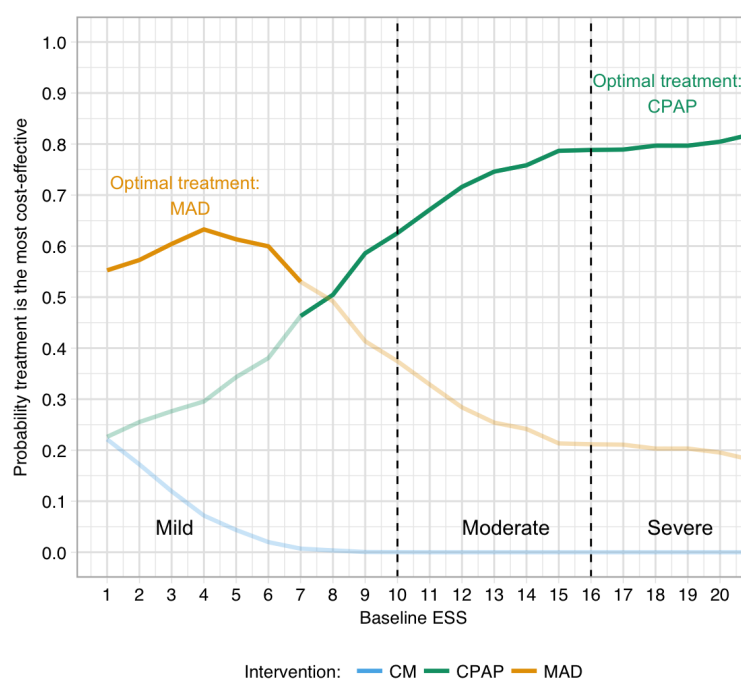


(a)Model 3: This model uses all AD and IPD and assumed $\gamma_A = \gamma_W$

CPAP.

Figure 4.6 presents the results of the case study CEA for each possible value of the ESS as the probability each intervention is the most cost-effective at a threshold of £20,000 per QALY gained. This presents further evidence in terms of both NMB and the probability the intervention is the most cost-effective that those with baseline ESS of less than seven should receive treatment with MADs and the rest CPAP.

Figure 4.6 Results of the case study cost-effectiveness analysis incorporating the results of the network meta-regression^(a) presented as the probability each intervention is most cost-effective, for each plausible baseline ESS value (shown in bold) and the intervention with the highest net monetary benefit at a cost-effectiveness threshold of £20,000 per QALY gained



(a)Model 3 - This model uses all AD and IPD and assumed $\gamma_A = \gamma_W$

4.5.2 Potential stratification of the optimal treatment decision on two covariates: Baseline ESS and BMI

Bayesian model-based network meta-regression results

Summaries of the posterior distributions for the network meta-regression parameters for each of the three models are shown along with their deviance statistics in Table 4.2.

When only AD is used (Model 1) and when $\gamma_W^j \neq \gamma_A^j$ (Model 2), the point estimates of γ_A^{MAD} indicate an increase of one unit of trial-level BMI or ESS is associated with increased ESS after treatment with MAD, although with wide credible intervals. This appears clinically incorrect and inconsistent with Table 4.1. There is some weak evidence that $\gamma_A^j \neq \gamma_W^j$ with the median values for γ_{A1}^{MAD} and γ_{W1}^{MAD} having opposite signs (also for γ_{A2}^{MAD} and γ_{W2}^{MAD}). However, the credible intervals for γ_{A1}^{MAD} and γ_{W1}^{MAD} overlap, and for γ_{A2} and γ_{W2} too. The results of the network meta-regressions which considered stratification on one covariate (Section 4.5.1) indicate a one unit increase in ESS reduces the ESS after treatment with MAD. When treated with CPAP, all models produced similar results, as expected. The network meta-regressions show similar results to the results of the network meta-regressions considering stratification on one covariate (Table 4.1) with BMI and the interaction of BMI and ESS having little impact on the results.

As with the network meta-regressions considering stratification on one covariate, the DIC can be compared between models 2 and 3, where a non-substantial difference was found. Similar to Section 4.5.1, the posterior estimates from Model 3 are used in the case study CEA, despite the weak evidence against this discussed above, as it provides more precise estimates by using between trial estimates of γ_A^{MAD} under the assumption of $\gamma_A^j = \gamma_W^j$ to estimate γ_W^{MAD} .

The impact of including the results from the network meta-regression on the case study cost-effectiveness analysis

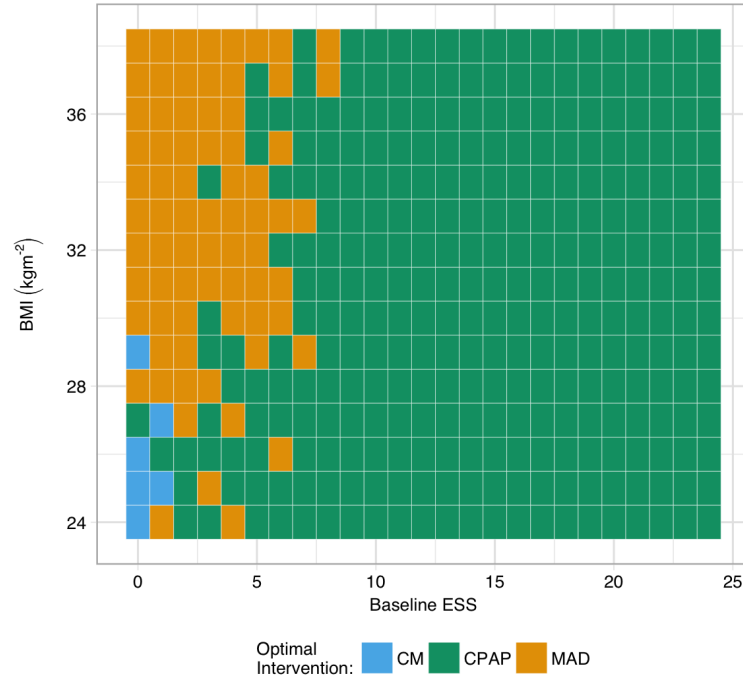
As in Section 4.5.1, the posterior estimates from Model 3 have been used in the case study CEA from Chapter 3. In this case, our interest is in stratifying the optimal treatment decision on the basis of two covariates - the ESS and the BMI. Figure 4.7 presents the intervention with the greatest INB at a threshold of £20,000 per QALY gained, for a range of ESS and BMI values.

Table 4.2 The posterior median, 95% credible intervals and deviance statistics for the coefficients of the network meta-regression models on the impact of baseline ESS and BMI on the change in ESS using aggregate^(a) and individual participant^(b) data with 100,000 Monte Carlo simulations and pooled effects as the means from the posterior distributions

	Model 1 AD	Model 2 AD + IPD $\gamma_A \neq \gamma_W$	Model 3 AD + IPD $\gamma_A = \gamma_W$
MAD^(c,d)			
ρ_{MAD}	-1.42 (-2.63, -0.25)	-1.54 (-2.68, -0.45)	-1.84 (-2.65, -1.07)
γ_{A1}^{MAD}	0.28 (-0.54, 1.10)	0.35 (-0.42, 1.11)	-0.15 (-0.46, 0.15)
γ_{A2}^{MAD}	0.20 (-0.35, 0.74)	0.21 (-0.34, 0.75)	-0.00 (-0.18, 0.18)
Γ_A^{MAD}	0.18 (-0.20, 0.56)	0.25 (-0.09, 0.59)	0.01 (-0.05, 0.08)
γ_{W1}^{MAD}		-0.38 (-1.07, 0.32)	
γ_{W2}^{MAD}		-0.25 (-0.82, 0.34)	
Γ_W^{MAD}		-0.24 (-0.59, -0.11)	
CPAP^(c,d)			
ρ_{CPAP}	-2.39 (-3.03, -1.82)	-2.42 (-3.06, -1.86)	-2.44 (-3.06, -1.89)
γ_{A1}^{CPAP}	-0.31 (-0.55, -0.12)	-0.32 (-0.55, -0.12)	-0.33 (-0.55, -0.14)
γ_{A2}^{CPAP}	-0.01 (-0.29, 0.26)	-0.02 (-0.30, 0.25)	-0.02 (-0.27, 0.24)
Γ_A^{CPAP}	-0.00 (-0.12, 0.12)	0.00 (-0.11, 0.12)	0.00 (-0.11, 0.11)
Deviance Statistics^(e)			
\bar{D}	141.9	1,031	1,032
pD	21.1	33.1	30.1
DIC	163.0	1,064	1,062

(a) AD; (b) IPD; (c) ϕ_j = The pooled treatment effect for the mean baseline ESS; (d) γ^j = The effect of a one unit increase in individual (W) or average (A) baseline ESS on the treatment effect with intervention j ; (e) see Chapter 3 for further information on these quantities.

Figure 4.7 Results of the case study cost-effectiveness analysis using the results of the network meta-regression^(a) presented in terms of the intervention with the highest net monetary benefit for all baseline ESS and plausible BMI combinations at a cost-effectiveness threshold of £20,000 per QALY gained



(a)Model 3 - This model uses all AD and IPD and assumed $\gamma_A = \gamma_W$

Compared to Figure 4.5, the impact of BMI on the optimal treatment decision is small for a given baseline ESS value. For those with lower BMI the baseline ESS required to change the optimal treatment decision from MAD to CPAP is lower, suggesting the lower the BMI the more likely it is that MAD would be the optimal intervention. This is supported by Table 4.2 which shows small values for γ_{A2}^j (the impact of a unit change in BMI) and in Γ_A^j (the impact of a one unit change in the study-level interaction between study-level baseline ESS and BMI). However, this is all based on a small amount of data (Figures 4.1 - 4.3).

4.6 Discussion

Limited data were found to help guide stratification on the basis of non-binary measures of disease severity (ESS and BMI) on the difference in ESS with interventions. The literature review (Section 4.2) found sparse data relating to treatment with MADs for patients with

OSAHS.

The chosen network meta-regression models for exploring the impact of baseline ESS and BMI on the difference in ESS between treatments assumed consistent within and between trial treatment covariate interactions, i.e. $\gamma_A^j = \gamma_W^j$ for all interventions j . Using all available evidence indicated an increased treatment effect for patients with greater levels of sleepiness. That is, both interventions reduce the ESS more for patients with higher initial ESS.

The network meta-regression models all showed greater uncertainty around the parameters pertaining to MAD, which was expected due to the lack of data. However, the addition of the IPD, which was all from studies comparing treatment with MAD and CM, reduces the uncertainty around the comparison of MAD and CM.

Incorporating the network-meta regression results into the case study CEA model, when considering stratification of the optimal treatment on one covariate, provided evidence that it may be beneficial to stratify the optimal treatment decision. On the basis of the case study CEA, those with an ESS of seven or less at baseline would be recommended treatment with MAD and the rest of the population recommended treatment with CPAP. Chapter 5 assesses, in detail, the population-level value of this stratification by considering the distribution of the stratifiers in the population, the level of adherence to the stratified regime and the costs of implementing a stratified regime. For potential stratification of the treatment decision by BMI and ESS, there appears to be little additional effect of stratifying on BMI. The extra value to the population created by stratifying on these two covariates is also explored in Chapter 5.

4.6.1 Data limitations

The studies in the literature review were the same as those used in Sharples et al. (2014, 2015) which were accessible and provided sufficient information [198, 199]. The one paper retracted since the publication of Sharples et al. (2014, 2015) has been removed from the analysis [196, 198, 199]. All studies were RCTs, being a mix of cross-over and parallel trials. Sharples et al. (2014) carried out an assessment of the quality of the studies but no further checks have been made [198]. This review is not comprehensive, but is believed to be representative of the available studies.

Data from two trials provided easily accessible IPD that could be used in the meta-regression. The authors of Sharples et al. (2014) kindly gave access to the underlying IPD from

TOMADO [198]. Hans et al. (1997) reported sufficient IPD in their paper for inclusion in the meta-regression [89]. While an IPD meta-regression is considered the ‘gold-standard’ it was decided not to approach authors of other studies to request IPD. Requesting and waiting for additional IPD is a time consuming and lengthy process which can yield little or no extra information. The aim of this work was to identify what decisions could be made with currently available data. Access to IPD from one study and finding further IPD from the publication of a smaller trial is a feasible situation. NICE recommend stratification of the optimal decision should be considered, meaning methodology for this should be able to be implemented within the, already tight, timescale for HTA reports and submissions - 12 weeks for a single technology appraisal and six months for a multiple technology appraisal [146]. It would be useful to formally assess the expected incremental gain from additional IPD by contacting researchers who publish AD. The value of collecting more IPD for a variety of different study designs, interventions and population sizes is explored in Chapter 5.

The RCTs in the network meta-regression are a mix of parallel and crossover trials. No formal adjustments have been made for this. Both trials providing IPD were cross-over trials. Here, the data has been treated as though from parallel trials, with each individual providing two independent pieces of information - the difference from the baseline ESS after treatment with CM and with MAD. This is a common assumption [65]. Both trials implemented a washout period between the periods of data collection. This involved participants not receiving any treatment to eliminate the treatment effect from the previous period. Due to the nature of the interventions, there is not thought to be a long lasting treatment effect. Therefore, treating the IPD as though from a parallel trial appears reasonable. The impact of the difference between cross-over and parallel trials in the AD trials has not been considered for the same reasons as for IPD: most cross-over trials implemented a washout period.

Issues around the recording of the ESS should be considered. The ESS is a self-reported questionnaire based on how likely you are to have fallen asleep in various everyday situations over the last few days. This indicates there may be reporting bias (Appendix A). However, the ESS is a widely used and validated questionnaire. Issues regarding regression to the mean and the short term nature of the trials used in the meta-regression may also occur with repeated measures of the ESS. These will occur with any patient reported outcome. Ideally, there would be more information on the differences between treatments over a longer time period. However, the value of collecting this data would be difficult to quantify. Patel et al. (2017) found the minimal clinically important difference for treatment with CPAP was a two

unit change in ESS [160]. Our models found this for the median pooled estimates when the baseline ESS was greater than 13 for MAD or greater than ten for treatment with CPAP.

4.6.2 Methodological issues

The impact of baseline ESS and BMI on the difference in ESS between treatment arms has been explored. The change in ESS with treatment is one of the two treatment effects considered in the case study CEA [198]. BMI is not a parameter in the CEA, however it is linked to patients' risk of OSAHS and CVEs. These may not be the most appropriate stratifiers to consider. A second treatment effect was used in the CEA. The impact of treatment on SBP was used to reflect the relationship between the interventions and the risk of CVEs presented through the use of the Framingham equations (Appendix B) [7]. Some preliminary work explored stratifying the optimal treatment decision on SBP. However, the data from the literature review on the relationship between baseline SBP and SBP after treatment with MAD and CPAP was scarce and not easily tractable. A number of different measures of SBP can be recorded - the average SBP over the day, over the night and over 24 hours. The relationship between these SBP values is not clear and there was insufficient data to consider stratification using data for just one of these definitions. Value of information methods could be used to find the value of collecting additional evidence on stratifying decisions on SBP. It is important that exploration of stratification is only carried out on parameters where there is a plausible reason on why the treatment may differ between participants and where there is sufficient data. Chapter 5 looks at the value of collecting information to guide stratification.

In the network meta-regressions the relationship between the covariate(s) and the difference(s) between treatments in terms of ESS has been assumed linear. In reality, other relationship structures may be preferable, for example a quadratic relationship. By inspection of the raw data from the studies in the literature review (Section 4.2) the relationship appears approximately linear. Adding additional terms to the regression or an alternative meta-regression structure, such as assuming a non-parametric relationship adds extra complexity which due to a lack of good data is likely to lead to issues with over-fitting and excessive uncertainty. Therefore, a linear model has been used as it is the simplest model that seems reasonable. These issues of excessive uncertainty were seen when considering stratification on two covariates, where the addition of the second covariate led to increased uncertainty around the treatment effect (Table 4.2).

In this work Φ^2 , the variance of the underlying individual participant observations of the change in ESS with the intervention over time, is estimated from the IPD alone. The study-

level variance of the change of ESS between treatment arms using AD ($\hat{\sigma}_{ij}^2$) was estimated using the reported individual study-level standard errors. This uses all available information and is used in Riley et al. (2008) and Thom et al. (2015) [182, 225]. However, it can lead to inconsistencies between the AD and IPD models. The alternative model structure which estimates Φ^2 from IPD and sets the standard error for the AD to be function of Φ^2 and the study size is consistent but does not use the published standard errors from the AD. A third option is to estimate Φ^2 from a combination of IPD and AD. This would use all available data and provide consistent models between the IPD and AD. However, estimating Φ^2 from the published standard errors from the AD may be complex, with potential little gain in accuracy over the approach used in this thesis.

Many of the studies used in the network meta-regression have similar baseline ESS, between 10 and 15. No study has a mean baseline ESS outside of the range 6 to 17. As the ESS can take values from zero to 24, extrapolation would be needed to assess the treatment effect for extreme values of the baseline ESS. The results of the network meta-regression reflect the lack of data at these extremes, due to the increased width of the credible intervals. However, it may be a strong assumption that the relationship is linear at these extremes, since this could lead to ‘impossible’ ESS values after treatment: i.e. that the treatment effect takes the ESS outside the range zero to 24, (discussed further below).

The ESS is bounded between zero and 24, therefore allowing a linear relationship for the network meta-regression may produce ‘impossible’ values at the extremes of the data range. The results indicate this is not of great consequence. At a baseline ESS of zero, 4.9% of the posterior samples give an ESS after treatment with CPAP outside the zero to 24 range. This compares to 43.3% of the posterior samples looking at treatment with MAD. It is worth noting that, at very low values of baseline ESS it is not infeasible that treatment with MAD or CPAP would increase a participants’ ESS score. While this seems counter-intuitive, from a practical point of view if a patient is not suffering in terms of sleepiness using an intervention may increase their daytime sleepiness.

In the network meta-regression the treatment effect has been assumed additive between intervention comparisons. That is:

$$\mu_{MAD-CPAP} = \mu_{MAD} - \mu_{CPAP}$$

Due to the lack of data on the MAD-CM and the MAD-CPAP treatment comparison this was difficult to test. However, it appears to be a reasonable assumption to make.

The main decision when choosing the network meta-regression model for use in the CEA concerns whether it is appropriate to assume the within and between study effects are the same (i.e. $\gamma_A^j = \gamma_W^j$). The credible interval for γ_W^j lies within that for γ_A^j and we can think of no obvious confounding factors that would lead to ecological bias. Therefore, the within and between study effects have been assumed equal, allowing the AD to be used to estimate the within study effects.

4.6.3 Future research priorities

This chapter indicates that individual patients should be recommended different optimal treatments on the basis of their ESS at diagnosis. There was little additional information to support the idea that stratification should be performed on BMI alongside baseline ESS. However, the results from this chapter are not sufficient to answer the question: should stratification on the ESS and BMI be carried out in practice? To answer this, the population distribution of ESS scores, the value of stratification, and any costs associated with implementing a stratified treatment decision need to be considered. This is explored further in Chapter 5.

Additional data on populations lying at the extremes of the ESS and BMI distributions, where there is a lack of data, could help to assess the assumption of a linear relationship between covariates and treatment effect. The expected value of collecting this data could be estimated by extending the meta-regression model to include the extra terms of potential interest, such as a quadratic term, with weakly informative priors. The results from this meta-regression could be used in the CEA and value of information measures estimated for these additional covariate parameters. This follows the principle of the ‘discrepancy’ approach by Strong et al. (2014a) [213].

4.7 Conclusion

There is limited data available on the impact of the baseline ESS and BMI on the change in ESS with treatment with MAD or CPAP for patients with OSAHS. By using Bayesian network meta-regression methods making use of all available data including AD and IPD, the impact of baseline ESS, and BMI on the change in ESS under treatment has been modelled, and the associated uncertainty quantified. By applying the results of this Bayesian network

meta-regression to the case study CEA it has been possible to identify those individuals in the population who would benefit from alternative interventions.

Chapter 5 extends upon the results from this chapter to explore the population-level stratification decision and assesses what further information would be useful to collect to guide stratification.

Chapter 5

The value of stratification and collecting further information to guide stratification

Chapter 4 presented methods for evidence synthesis to inform stratified treatment decisions for non-binary covariates at an individual patient-level. It found for the case study CEA the treatment with the highest NMB depended on the patients' baseline ESS and to a lesser extent BMI. This chapter extends upon this work to assess whether a stratified regime should be implemented at a population-level by calculating the health economic value of stratification. It considers the population distribution of the strata, the level of adherence by physicians and patients to the stratified regime, and the costs associated with implementing a stratified treatment regime. Once a decision on stratification has been made there may be value in collecting further information to guide stratification. Methods for prioritising further research and for efficiently computing value of information quantities for stratified treatment decisions are presented. All of this work is applied to the case study CEA.

5.1 Introduction

Under current information the optimal intervention, in terms of cost-effectiveness, may differ between groups of the population. However, the presence of different optimal interventions for subgroups is not a sufficient condition for implementing a stratified regime. For implementation to be considered worthwhile in an economic context, factors such as the costs of stratification, the distribution of the strata in the population, and the extent it is expected

physicians and patients might adhere to the stratified regime need to be considered.

The work in Chapter 4 indicates, on an individual patient basis, value in stratification for the case study CEA. However, a population-level perspective is needed to assess whether the regime should be implemented. Health economic decisions, such as those made by NICE, consider the value of a strategy to the population. Therefore, factors such as those mentioned above need to be considered. In particular, while individual patients with specific values of stratifiers may be recommended different interventions, if the proportion of the population who would be given an intervention considered suboptimal under a non-stratified regime is small then the related population-level value of stratification could be small.

In such a scenario the monetary costs of stratification become even more important. The process of allocating people to their correct stratum has a cost. If this is greater than the population-level value of stratification there is no health economic value of stratification. The extreme of this is the idea that everyone should be treated compared to the scenario where people are then tested and allocated treatment on the basis of the test results in an outbreak scenario.

The idea that a stratified regime may not be adhered to completely by physicians (or patients) is a key concept as it can significantly impact whether a stratified treatment decision should be implemented [47]. The presence of a set of pre-defined values at which patients' prescribed intervention may change could lead to patients close to these boundary values being prescribed a suboptimal intervention. This 'blurring' of the strata allocated to each intervention can reduce or eliminate the value of the stratified regime.

This chapter extends upon the work in Chapter 4. There, evidence was found for the case study CEA that stratification of the optimal intervention may be useful. There are two main aims to this chapter. The first is to present a generalised version of currently available methods to assess whether, given the strata, the stratification regime should be implemented. The key components are the costs of stratification and the distribution of the strata in the population. Methodology is presented to allow for calculating the value of stratification for discrete and continuous covariates. Additionally, the impact of different population distributions for the stratifying variables and the impact of suboptimal treatments being given is investigated in the case study. This theory is outlined in Section 5.2 and applied to the results of the case study CEA (Chapter 4) in Section 5.2.2.

After deciding whether to implement a stratified treatment regime under current information, the next logical step is to assess whether further information should be collected to reduce uncertainty relating to stratification. The concepts of EVPPI and EVSI from Section 1.1.3 and Chapter 2 are revisited in the context of stratified decision making (Section 5.3). Extensions of current methods are developed to allow for efficient calculation of EVPPI and EVSI for stratified decision making using a single non-parametric regression. Applications to the case study CEA illustrate these methods. Section 5.4 extends and answers some of the discussion points from Chapter 4 and additionally details limitations of this work and further research priorities.

5.2 The health economic value of stratification

5.2.1 Background and theory

Chapter 4 found the optimal treatment for the case study CEA differed for patients according to their baseline ESS and BMI values. However, it did not quantify whether population-level treatment allocation should depend on these stratifiers. Both clinical and cost-effectiveness criteria need to be considered. Chapter 4 presents the clinical value of stratification: through the Bayesian network meta-regression combining IPD and AD it found the treatment effect differs between interventions and patient characteristics. The case study CEA used the results from the network meta-regression and found those with a baseline ESS of less than seven should be prescribed MADs and the remainder of the population prescribed CPAP. Additionally, for potential stratification of the treatment decision by BMI and ESS there appears to be little additional effect of stratifying on BMI. However, under a NHS type healthcare system this is not sufficient for implementation. Stratification needs to provide additional value to the population as a whole.

The remainder of this section presents a number of concepts relating to the value of stratification under current information. Even if there are individual patient benefits to stratification, stratification should only be implemented if the population-level costs of stratification are less than its population benefit. There are also issues around non-adherence to stratification regimes due to a concept named leakage.

The value of stratifying the optimal treatment decision under current information

Espinoza et al. (2014) presented work on the value of heterogeneity for categorical subgroups (normally binary) of the population [67]. Here this work is used to present a more general

framework for stratification on discrete and continuous covariates [67]. Taking a UK perspective, the aim of the NHS is to maximise the health of the population subject to a financial budget [145]. This work explores the value of taking into account stratification variable(s), \mathbf{X} , in treatment allocation. Let $j \in \{0, 1\}$ index the set of interventions to which each strata could be allocated. As previously, let $\boldsymbol{\theta}$ be the set of uncertain parameters in the CEA. Assume the allowed values of \mathbf{X} can be partitioned into mutually exclusive categories (for example, taking integer values only). \mathbf{X} could be a single stratifier or a vector of stratifiers. Assume, for now, that \mathbf{X} is a single stratifier.

The cost-effectiveness problem in terms of the EIB between interventions 0 and 1 for a given threshold, λ is:

$$\begin{aligned} EIB_{0-1}(\mathbf{X}) &= E_{\boldsymbol{\theta}}[NB(j=0, \boldsymbol{\theta}, \mathbf{X})] - E_{\boldsymbol{\theta}}[NB(j=1, \boldsymbol{\theta}, \mathbf{X})] \\ &= ENB(j=0, \mathbf{X}) - ENB(j=1, \mathbf{X}) \end{aligned}$$

where $NB(j, \boldsymbol{\theta}, \mathbf{X})$ is the NMB for the population with stratification variable \mathbf{X} and intervention j :

$$NB(j, \boldsymbol{\theta}, \mathbf{X}) = \lambda \bar{e}_{j,x} - \bar{c}_{j,x} \quad (5.1)$$

where $\bar{e}_{j,x} = E_{X=x}(e_j)$ and $\bar{c}_{j,x} = E_{X=x}(c_j)$.

If $EIB_{0-1}(\mathbf{X}) > 0$, for a stratum with stratifier \mathbf{X} , then intervention 0 is more cost-effective. Similarly, if $EIB_{0-1}(\mathbf{X}) < 0$ then intervention 1 is more cost-effective.

$EIB_{0-1}(\mathbf{X})$ can be calculated for each value of \mathbf{X} . The change in optimal treatment occurs at the x where $EIB_{0-1}(\mathbf{X})$ changes sign. Let x^* be the largest x where $EIB_{0-1}(\mathbf{X}) > 0$, so if $EIB_{0-1}(\mathbf{X})$ is monotonic in \mathbf{X} :

$$\begin{array}{ll} x \leq x^* & j=0 \text{ is optimal} \\ x > x^* & j=1 \text{ is optimal} \end{array}$$

The Expected Net Benefit (ENB) for the population is maximised when each stratum receives their optimal intervention. The maximal ENB under stratification by \mathbf{X} is:

$$ENB^{(\mathbf{X})} = \sum_{x \leq x^*} ENB(j=0, x) w_x + \sum_{x > x^*} ENB(j=1, x) w_x \quad (5.2)$$

where w_x is the proportion of the population where $\mathbf{X} = x$; $\sum_x w_x = 1$ and x^* is the value of $x \in \mathbf{X}$ where the optimal treatment changes.

If the ENB is not monotonic in \mathbf{X} then j may be optimal for disjoint subgroups. For example, suppose \mathbf{X} can take values $\{x_1, x_2, x_3, x_4\}$ where $x_1 < x_2 < x_3 < x_4$, and $j = 0$ is optimal for $\{x_1, x_3, x_4\}$ and $j = 1$ optimal for $\{x_2\}$. In this example, there are two points where the optimal treatment changes: x_2 and x_3 .

Let $\{\mathbf{X}_j\}$ be the subset of the values of \mathbf{X} where intervention j is optimal, $j = 0, 1$. Equation 5.2 can be re-expressed as:

$$ENB^{(\mathbf{X})} = \sum_{x \in \{\mathbf{X}_0\}} ENB(j=0, x) w_x + \sum_{x \in \{\mathbf{X}_1\}} ENB(j=1, x) w_x \quad (5.3)$$

Similarly, if there are more than two interventions, each optimal for a different subset of strata $\{\mathbf{X}_j\}$, $j = 0, \dots, J$, then Equation 5.3 can be extended:

$$ENB^{(\mathbf{X})} = \sum_j \sum_{x \in \{\mathbf{X}_j\}} ENB(j, x) w_x \quad (5.4)$$

If \mathbf{X} is a vector corresponding to stratification on n covariates, then Equation 5.4 can be used with each $\{\mathbf{X}_j\}$ corresponds to the n -tuples where j is optimal.

The structure above assumes \mathbf{X} can take one of a finite set of mutually exclusive values. If \mathbf{X} is truly continuous (for example, weight), then:

$$ENB^{(\mathbf{X})} = \sum_j \int_{x \in \{\mathbf{X}_j\}} ENB(j, x) f(x) dx \quad (5.5)$$

where $f(x)$ is the probability density function of \mathbf{X} . In Equation 5.5, there is a sum over j as the set of interventions remains discrete and mutually exclusive. While stratifiers can be continuous, they are often measured or recorded on a discrete scale, in which case \mathbf{X} can be considered as discrete (for example - weight may be recorded to the nearest kilogramme). Additionally, policy-makers may prefer to define specific values which consider the accuracy the stratifier could be recorded or measured, in which case Equation 5.4 can be used.

Equation 5.4 presents the ENB for the population when a stratified treatment regime is implemented. A number of papers have explored the population-level value of stratification [19, 47, 67]. This is an important concept and is used to decide whether to implement a stratified treatment regime.

The static value of heterogeneity is defined by Espinoza et al. (2014) (Coyle et al. (2003) present a similar argument), as the value of stratifying the optimal treatment decision under current information equivalent to the difference between the population-level ENBs with and without stratification [47, 67]:

$$\sum_j \sum_{x \in \{X_j\}} [ENB(j, x)w_x] - E_{\theta} \left[\max_j NB(j, \theta) \right]$$

As the set of potential values that \mathbf{X} can take increases, either in terms of granularity of a single stratifier or by the number of stratifiers, the number of strata increases. Taking this to its limit gives the case where individualised care is offered [18].

Non-adherence to optimal treatment allocation

Coyle et al. (2003) present the idea that if the treatment decision is stratified, physicians may not strictly adhere to the stratification regime [47]. This is a similar, but different, concept to the uptake rate for an intervention (Chapter 2) where uptake is defined as the proportion of the eligible population who use their intervention [86]. Non-adherence to a stratified regime is perhaps a more pertinent issue in the case of continuous or discrete stratifiers as opposed to a binary stratifier. Those individuals whose covariate values are close to x^* may be more likely to be prescribed a suboptimal treatment, as opposed to a binary stratifier where it should be clear to which stratum each patient belongs to.

Coyle et al. (2003) define ‘leakage’ as the proportion of patients prescribed a suboptimal treatment [47]. Let L_x be the proportion of patients with stratifier value(s) $x \in \mathbf{X}$ who receive a suboptimal treatment. If $j = \{0, 1\}$, then the ENB taking into account leakage, $ENB^{(\mathbf{X})}|L$ is:

$$ENB^{(\mathbf{X})}|L = \sum_j \sum_{x \in \{X_j\}} [ENB(j, x)(1 - L_x) + ENB(\bar{j}, x)L_x]$$

where \bar{j} is the suboptimal treatment for the population in strata x .

If there are more than two interventions then the situation is more complex. Leakage could, theoretically, occur into any other intervention. Let $\tilde{L}_{j,x}$ be the proportion of people with set of stratifiers x receiving treatment j , then:

$$ENB^{(\mathbf{X})}|L = \sum_j \sum_{x \in \{X_j\}} ENB(j, x) \tilde{L}_{j,x}$$

$ENB^{(\mathbf{X})}|L$ can be less than the ENB under no stratification when those receiving the suboptimal treatment lose more NMB than they would gain from stratification negating any extra value found by stratification. Coyle et al. (2003) also noted leakage is an endogenous product of the presence of leakage [47]. That is, should physicians know an element of leakage has been incorporated in the stratification decision they may be more likely to prescribe a suboptimal intervention, causing the true rate of leakage to rise.

Leakage may not be due to the physicians' actions alone. It could also occur when patient preferences are considered. Patients with an x close to x^* may be offered a choice of interventions. A patient given an opportunity to express a treatment preference may choose a treatment that, in terms of effectiveness, is suboptimal. However, a patient who has chosen their intervention may have an increased adherence to their intervention. The adherence rate combined with the effectiveness of the intervention of choice may lead to a higher NMB in practice. Chapter 3 showed how adherence to interventions can impact the optimal treatment decision.

It is also worth noting that in addition to leakage the implementation of a stratified intervention regime may not be perfect. This could be due to physician preference for a specific intervention. There may also be barriers to implementation due to extra time being needed to implement the stratification decision or hospitals not having access to an intervention or the equipment/tests needed to allocate patients to subgroups. There are a number of papers that explore this issue further [8, 73].

Costs associated with stratifying the optimal treatment decision

There are two types of costs relating to stratification: the additional costs required for research on how costs and effects vary between potential strata and costs associated with measuring stratifiers. This thesis explores the first in Section 5.3 and this section discusses the second.

In stratifying a treatment decision there may be costs associated with measuring \mathbf{X} for individuals. Phelps (2014) believes these costs could be more important than the choice of

stratification policies in terms of cost-effectiveness [164].

Patients can be allocated to strata using a number of methods and/or tests, for example by conducting a blood test, a questionnaire, or a genomic test. Let C be the cost of allocating one patient to an intervention¹. There is value in stratifying the optimal treatment if:

$$ENB^{(X)}|L - C > 0$$

For example, if the stratifier is the result of a genomic test, C can be large. Therefore, the extra value due to stratification, in terms of ENB, needs to be greater than the cost of assigning patients to subgroups. Conversely, if the stratifier was a lower cost test, such as a patient completed questionnaire (such as the ESS), the resulting increase in ENB would need to be less for stratification of the treatment decision to be valuable.

Additional costs may arise if further information on subgroups is collected to determine how strata should be defined. Phelps (1997) outline four main reasons as to why the cost-effectiveness of interventions may change across a population [163]. Firstly, stratification will require more research into the proportion of the population in each stratum, producing additional costs. Section 5.2.2 provides an example of this. Further, treatment efficacy may depend on the subgroup. This relationship can sometimes be identified using existing data without the additional costs of obtaining new data by using meta-regression (Chapter 4). Phelps (2014) points out costs can differ between subgroups due to differing risks of co-morbidities which may not be reflected in the existing CEA model [164]. For example, in the case study CEA, an increased baseline ESS may impact the risk of a RTA. This would not be reflected by the subgroups specification introduced in Chapter 4. The final point made by Phelps (1997) is that utility estimates may differ between subgroups and should be estimated separately for each subgroup which generates additional research costs [163]. Methods for estimating the expected value of collecting information to guide stratification are discussed further in Section 5.3.

The expected value of stratifying the optimal treatment decision under further information

As in the case of no-stratification, once the maximum ENB for the population of interest has been calculated, the uncertainty around the stratified treatment decision can be expressed in terms of the value of collecting further information to guide stratification and for other

¹This is done on a per person basis as the NMB is on a per person basis

parameters.

In a stratified decision, for each x under current information the intervention j with maximum expected NMB is chosen:

$$\max_j E_{\boldsymbol{\theta}} [NB(j, \boldsymbol{\theta}, x)]$$

Under perfect information, the maximum NMB is expected to be:

$$\max_j NB(j, \boldsymbol{\theta}, x)$$

However, as $\boldsymbol{\theta}$ is not known with certainty, the expectation with respect to $\boldsymbol{\theta}$ needs to be taken. Therefore the EVPI for stratum x is [67]:

$$EVPI_x = E_{\boldsymbol{\theta}} \left[\max_j NB(j, \boldsymbol{\theta}, x) \right] - \max_j E_{\boldsymbol{\theta}} [NB(j, \boldsymbol{\theta}, x)]$$

where $EVPI_x$ is the per person EVPI for an individual with covariates x , similar to the EVPI (Sections 1.1.3 and 2.2). This is the upper bound for further research on the population in stratum x , assuming stratification of the treatment decision.

The average per person EVPI for a mixed population is [67]:

$$EVPI_{\mathbf{X}} = \sum_x EVPI_x w_x \quad (5.6)$$

that is, the average EVPI over all strata ($EVPI_{\mathbf{X}}$) is a weighted sum of the EVPI for each strata ($EVPI_x$) on the prevalence of the strata in the population, w_x . This assumes \mathbf{X} is discrete. If \mathbf{X} is continuous then this can be generalised by integrating over the pdf of \mathbf{X} .

As in a non-stratified analysis, population-level EVPI is more useful for prioritising future research than per person EVPI. Let $w'_x(t)$ be the prevalence of each x in the population at time t , then the average per person EVPI at time t is:

$$EVPI_{\mathbf{X}}(t) = \sum_x EVPI_x w'_x(t)$$

If the population at all t have the same distribution of \mathbf{X} as the current population, i.e. we do not expect the population distribution of \mathbf{X} to change with time, then $w_x \equiv w'_x(t) \forall t$.

For a time horizon T , with time indexed by $t = 1, \dots, T$, disease prevalence of I_0 , disease incidence rate at time t of I_t , population at risk of disease P , and discount rate d then the population-level EVPI for a stratified decision can be defined as in Section 2.2.2:

$$popEVPI_{\mathbf{X}} = P \times \left[EVPI_{\mathbf{X}}(0)I_0 + \sum_{t=1}^T \frac{I_t}{d^t} EVPI_{\mathbf{X}}(t) \right]$$

Espinoza et al. (2014), as part of their value of heterogeneity specification, define the Dynamic Value of Heterogeneity [67]. This is the expected value from collecting new data to reduce uncertainty on parameters explicitly related to how cost-effectiveness depends on the subgroups. This is defined to be the difference in the expected perfect information between a stratified decision and an non-stratified decision:

$$\sum_x E_{\boldsymbol{\theta}} \left[\max_j NB(j, \boldsymbol{\theta}, x) \right] - E_{\boldsymbol{\theta}} \left[\max_j NB(j, \boldsymbol{\theta}) \right]$$

The dynamic value of heterogeneity is the additional health, in monetary terms, the population would have if decision uncertainty was eliminated under stratification, compared to the non-stratified case. This is the value of resolving uncertainty on the parameters the stratified decision is dependent on, i.e. \mathbf{X} . If the dynamic value of heterogeneity equals the static value of heterogeneity further research would be preferable on those parameters that do not affect the stratification.

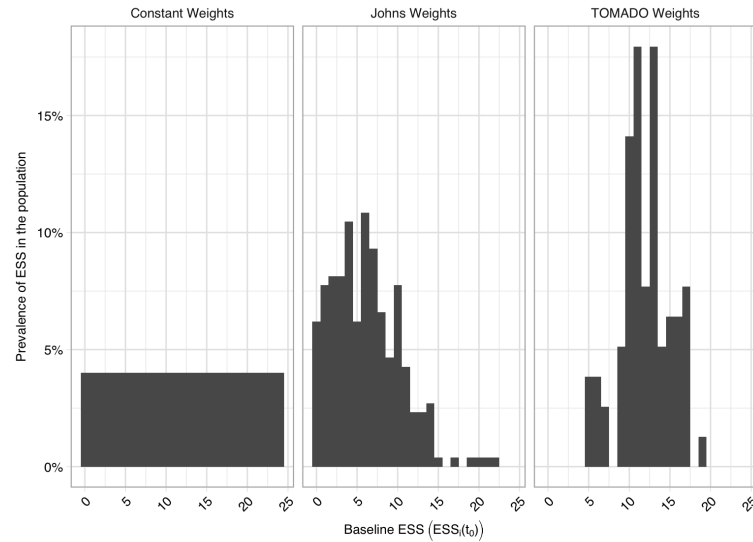
Basu and Meltzer (2007) developed the concept of Expected Value of Individualised Care (EVIC) motivated by the US healthcare system which defined as how much society should be willing to pay to make treatment decisions on an individualised basis [19]. It quantifies the benefit foregone by implementing a population-level decision as opposed to an individual-level decision [19, 231]. Individualised care is an extreme case of stratification where the set of stratifiers means each patient receives their optimal treatment.

The EVPI and the EVIC are complementary [231]. However, the EVPI presents the value of resolving uncertainty on the treatment decision at a population-level whereas EVIC does the same thing at an individual patient-level[231].

5.2.2 An illustration of the impact of the population distribution and leakage on the case study

This section explores the value of stratification for the case study CEA. A number of different distributions of baseline ESS and BMI over the population have been considered to calculate

Figure 5.1 Three alternative population distributions of ESS used to calculate the value of stratification for the case study cost-effectiveness analysis [114, 198]



the value of stratification assuming no leakage. The impact of leakage on the optimal treatment allocation under stratification is then applied to the case study.

The impact of the population distribution on stratification

Three distributions for baseline ESS in the population have been used to highlight the impact the population composition has on the value of stratification. These distributions are illustrated in Figure 5.1. As earlier, let w_x be the proportion of the population with a baseline ESS value of x . The first distribution, arbitrarily, assumes the proportion of people with each baseline ESS is constant over the whole range of ESS:

$$w_x = \frac{1}{25}, \quad x = \{0, 1, \dots, 24\}$$

The second distribution is taken from a study by Johns and Hocking (1997) looking at the daytime sleepiness, measured by the ESS, of Australian workers [114]. They used a population of 259 workers with a sleep disorder². The values for all participants with a sleep disorder have been taken as a proxy for the distribution of ESS in a population with OSAHS.

²A sleep disorder was defined using a broad criteria including poor quality of sleep; taking > 30 minutes to fall asleep; difficulty in falling asleep; waking up more than three times a night; moderate difficulty in going back to sleep; consulted doctor about sleep; takes sleep medication; snores; and stops breathing or makes choking noise. The final definition includes OSAHS where 30% of male and 14% of female responders were found to have this.

This study found a skew in the population towards a low ESS.

The third ESS population distribution is the baseline ESS distribution of the TOMADO population, the case study for this thesis (Section 1.3 and Appendix B) [198]. TOMADO selected patients with an inclusion criteria of baseline $ESS \geq 9$. Data was available for 78 participants ($5 \leq ESS \leq 19$)³. This population had a greater proportion with higher ESS values compared to the other populations under consideration.

Population weights from these three scenarios have been applied to the case study CEA from Section 4.4 to give the population average NMB for each intervention, $j = \{CM, MAD, CPAP\}$, under no stratification. In the unstratified analysis the baseline ESS is an input and the value of this is allocated on the basis of these weights in each PSA sample. However, when assessing the optimal intervention everyone receives the same intervention. The optimal ENB for a stratified decision was that patients with $ESS \leq 7$ should be treated with MADs and the remainder treated with CPAP (Figure 4.6). The NMB under stratification and the per person EVPI values for policies with and without stratification were calculated and used to calculate the static and dynamic values of heterogeneity (Table 5.1 [67]).

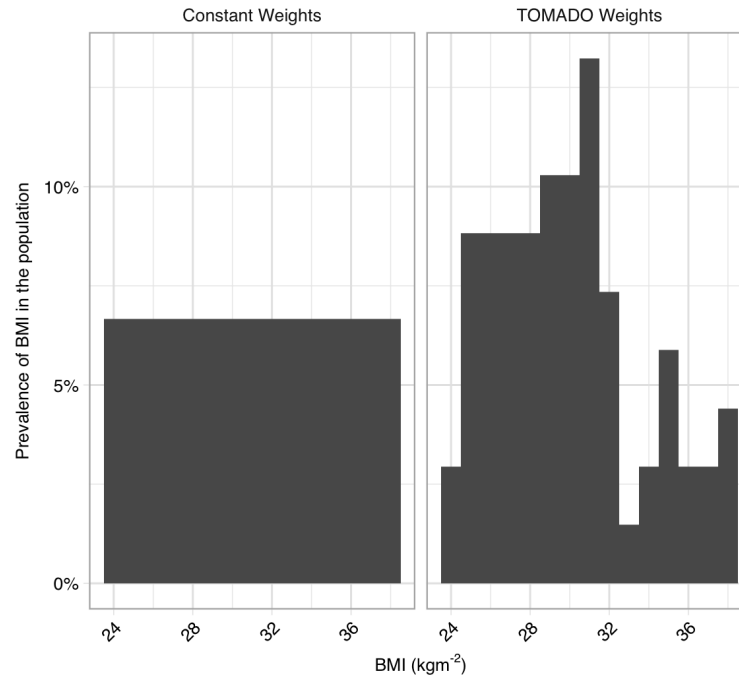
Changing the distribution of ESS in the population changes the optimal treatment under no stratification (Table 5.1). The value of heterogeneity, both in terms of static and dynamic value is sensitive to the distribution of the ESS in the population. In all cases, assuming no costs of stratification there appears to be value in stratification.

The stratifier in this work is the ESS which could be completed in one appointment with a consultant. In the case study CEA the mean cost of a consultant appointment was £106 (Table B.8). Using this as a proxy for the cost of stratification, if the population distribution was the same as the TOMADO population the costs of stratifying the optimal treatment decision would be £54 per person greater than the value of stratification. Therefore, stratification should not be implemented. However, there would still be value in stratification if the population had a constant distribution over the range of ESS or a distribution resembling that in Johns and Hocking (1997) [114].

For the constant weights and Johns weights of ESS in the population, the population-level value of static heterogeneity is £45 million and £100 million respectively, assuming the cost of the stratification is a single appointment with a consultant at £106 per patient. This shows

³A small number of participants were included despite their baseline $ESS < 9$ as at screening their $ESS \geq 9$.

Figure 5.2 Two alternative population distributions for BMI used to calculate the value of stratification under a two variable stratification policy for the case study cost-effectiveness analysis [198]



a considerable value to society through stratifying the optimal treatment decision. This is due to there being a larger proportion of the population having a baseline ESS less than seven compared to when using the population weights derived from the TOMADO. As it is this group of patients whose optimal treatment will change with stratification, this leads to a greater value of stratification.

For two variable stratification, a population distribution is required for both the ESS and the BMI, $\mathbf{X} = \{x_1, x_2\} = \{ESS, BMI\}$. Two alternative population distributions have been explored for BMI (Figure 5.2). The first, arbitrarily, assumes the proportion of people with each BMI is constant over the range $28\text{-}36\text{kgm}^{-2}$:

$$w_{x_2} = \frac{1}{9} \quad x = \{28, 29, \dots, 36\}$$

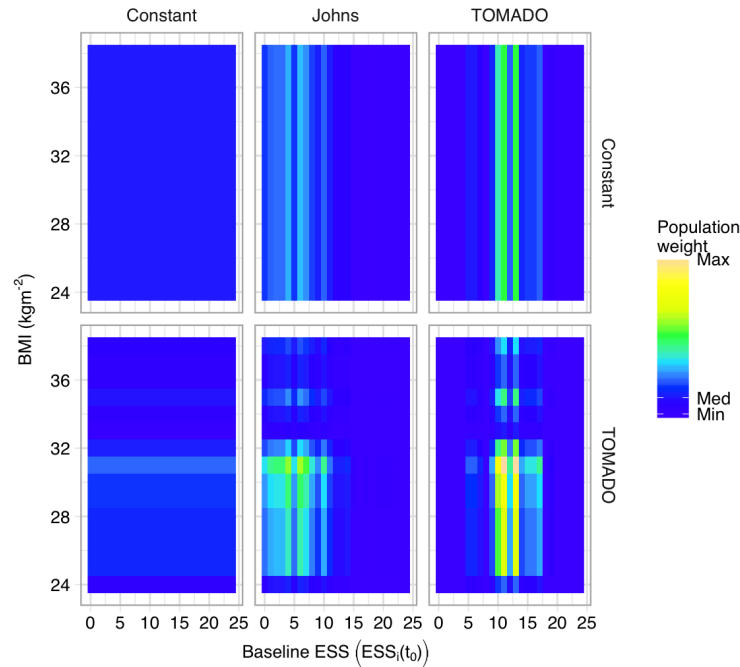
The second distribution considered is the empirical distribution of the 78 participants in the TOMADO, cut off at a BMI of 36kgm^{-2} . This population had a greater proportion of people with lower BMI.

Table 5.1 Results of the exploration of the value of heterogeneity for three different population distributions for stratification on ESS using the results of the cost-effectiveness analysis presented in Chapter 4

Value (£)		ESS Weight Distribution		
		TOMADO	Constant	Johns
NMB of unstratified policy ¹				
CM		280,706	280,626	292,839
MAD		284,757	284,741	296,126
CPAP		285,866	285,801	295,505
NMB under stratification ²	A	285,918	286,272	296,425
EVPI (£pp)				
No Stratification	C - bold	547	1,212	1,319
Stratification	B - A	559	1,217	1,993
NMB under Perfect Information				
No Stratification	C	286,413	287,013	297,445
Stratification	B	286,477	287,489	298,418
Individual Value of Heterogeneity ⁴ / Value of Stratification				
Static (£pp ³)	D = A - bold	52	471	920
Dynamic (£pp ³)	E = B - C	64	476	973
Cost of Stratification ⁵ (£pp)	F	106	106	106
Population-level Value of Heterogeneity ^{4,6,7} / Value of Stratification				
Static (£ million)	P × [D-F]	-7	45	100
Dynamic (£ million)	P × [E-F]	-8	46	107

¹maximum NMB highlighted in Bold; ²for the optimal intervention; ³per person ⁴see Section 5.2.1 and Espinoza et al. (2014) for explanation of these quantities [67]. ⁵assuming costs of stratification is the cost of one appointment with a consultant. ⁶assuming costs of stratification is the cost of one appointment with a consultant of £106pp and the population-level values estimated using the parameters from Section 2.2.2 ⁷P is the population-level scaling factor to transform the value from a per person to a population-level value (Section 2.2.2) taking into account the time horizon of the technology and the disease incidence and prevalence rates.

Figure 5.3 The weights defining the six population distributions of combined ESS and BMI values (assuming independence) used to calculate the value of stratification under a two covariate stratification policy for the case study cost-effectiveness analysis [114, 198]



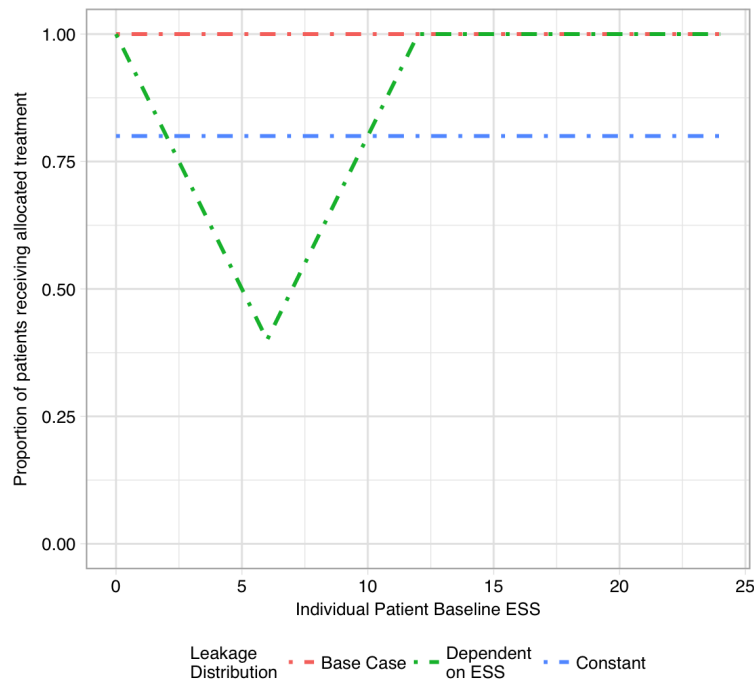
Based on the results in Chapter 4, the relationship between ESS and BMI is assumed to be independent which gives the joint ESS and BMI distributions presented in Figure 5.3:

$$w_{x_1 x_2} = w_{x_1} w_{x_2}$$

$$x_1 = \{0, 1, \dots, 24\} \quad x_2 = \{28, 29, \dots, 36\}$$

These weights are applied to the case study CEA in Section 4.4 to give the population-level values of stratification for each combination of weights for BMI and ESS (Table 5.2). The population distributions have less of an effect in this case. As in Table 5.1, if the ESS population distribution was as in Johns and Hocking (1997) then MAD would be the optimal intervention under no stratification, regardless of the population BMI distribution [114]. Similarly, assuming the population ESS distribution was constant or as in the TOMADO then CPAP would be the optimal treatment under no stratification [198]. There is value in stratification for all population distributions assuming no costs of stratification. However, if the costs of allocating a patient to a stratifier remained at £106 per person (assuming measuring the BMI does not require additional time or incur further costs) there would be no

Figure 5.4 The proportion of patients receiving their optimal treatment, as suggested by the case study cost-effectiveness analysis under two alternative leakage distributions compared to the base case



value in stratification when the ESS population distribution was similar to that in TOMADO regardless of the population distribution of BMI.

The impact of leakage on the value of stratification for the case study

For ease of calculation, this section assumes the distribution of the population is defined by the following fixed values for the ESS, $\mathbf{X} = x_1$:

$$w_{x_1} = \frac{1}{25}, \quad x_1 \in \{0, 1, \dots, 24\}$$

However, this method and application can easily be extended to different population distributions and multiple stratifiers. When considering stratification by ESS Figure 4.6 showed value in prescribing MADs to those with $ESS \leq 7$ and CPAP to those with $ESS > 7$. This implicitly assumes full implementation and no leakage.

Two hypothetical leakage distributions have been considered (Figure 5.4). The first, L_1 , assumes the rate of leakage is greatest at values of ESS with the most uncertainty around the

Table 5.2 Results of the exploration of the value of heterogeneity for six different population distributions for stratification on BMI and ESS using the results of the case study cost-effectiveness analysis in Chapter 4

ESS BMI Value (£)	Population distribution of stratifiers					
	Constant	Constant	TOMADO Constant	Johns Constant	Constant TOMADO	Johns TOMADO
NMB ¹						
CM	280,636		280,728	292,846	280,642	280,742
MAD	284,657		284,803	295,732	284,725	284,885
CPAP	286,730		286,786	295,635	286,700	286,757
NMB under stratification ²	287,137		286,811	296,433	287,072	286,780
EVPI (£pp)						
No Stratification	2,006		773	4,398	1,972	736
Stratification	2,001		753	3,249	1,900	704
NMB under perfect information / Value of Stratification						
No Stratification	288,138		287,559	299,682	288,672	287,493
Stratification	288,736		287,564	300,130	288,972	287,484
Individual-level value of heterogeneity ³ / Value of Stratification						
Static	407		25	701	372	23
Dynamic	402		5	448	300	-9

¹maximum NMB highlighted in Bold; ²for the optimal policy; ³see Section 5.2.1 and Espinoza et al. (2014) for explanation of these quantities [67].

optimal treatment decision, i.e. ESS is close to 6. A second distribution, L_2 , assumes 20% of patients receive a suboptimal treatment regardless of their ESS.

Without stratification the maximum ENB was £285,801 with the whole population receiving treatment with CPAP. When stratification is implemented fully (i.e. no leakage and everyone receives their optimal treatment) the ENB was £286,272 (Table 5.1). Under L_1 dependent on ESS the $ENB^{(X)}|L_1$ is £286,152. The static value of heterogeneity reduces by £120 per person from £471 per person to £351 per person.

Under a constant leakage distribution, L_2 , the $ENB^{(X)}|L_2$ is £285,872. The static value of heterogeneity reduces to £71 per person indicating no value of stratification should the cost of implementing the stratified policy be the £106 per person cost of an appointment with a consultant. This brief exploration has shown that the extent of leakage may change the decision on the implementation of a stratified policy.

5.3 The value of collecting further information to guide stratified decision making

When stratifying the optimal treatment decision it is important to assess whether further information on how cost-effectiveness varies between subgroups may be valuable. In particular, if there is value in collecting information it would be useful to identify the populations for whom the value is greatest.

In Chapter 2 a number of methods for the calculation of the EVPPI and the EVSI were presented. The methods by Strong et al. (2014, 2015) [215, 216] are extended in this section to estimate the value of information for a stratified decision.

5.3.1 The Expected Value of Perfect Partial Information for parameters related to stratification

When considering stratification on \mathbf{X} the EVPPI can be calculated for each stratum and aggregated to gain the average per person EVPPI for the stratified population, using the methods of Section 2.3.2. Alternatively, it is likely to be more efficient to calculate the EVPPI for all strata simultaneously (dependent on how many strata there are).

A multiple calculation method for estimating the Expected Value of Perfect Partial Information relating to stratified treatment decisions

Formally, let $\widehat{EVPPI}_x(\phi)$ be the estimate of EVPPI calculated for a set of parameters $\phi \subseteq \theta$, using the methods of Strong et al. (2014) for a stratum value x [215]. Let w_x be the proportion of the population in stratum x . The estimated average per person EVPPI for a mixed population, $\widehat{EVPPI}^{(X)}(\phi)$ is:

$$\widehat{EVPPI}^{(X)}(\phi) = \sum_x \widehat{EVPPI}_x(\phi) w_x$$

as in Equation 5.6, with EVPI replaced by EVPPI. These estimates of $\widehat{EVPPI}_x(\phi)$ and w_x can be used to calculate the population-level EVPPI for a stratified decision using the methods in Section 2.3.2.

Let s_x be the standard error associated with $\widehat{EVPPI}_x(\phi)$. The standard error for $\widehat{EVPPI}^{(X)}(\phi)$, s , can be calculated by:

$$s = \sqrt{\sum_x w_x^2 s_x^2}$$

Thus $\widehat{EVPPI}^{(X)}(\phi)$ and its standard error can be estimated by calculating \widehat{EVPPI}_x and s_x^2 for each x separately, using the methods in Strong et al. (2014) (Section 2.3.2). This requires many more PSA samples for the same level of precision as a non-stratified EVPPI estimate which results in either an increase in computational time for the CEA and EVPPI or a decrease in precision.

A single calculation method for estimating the Expected Value of Perfect Partial Information relating to stratified treatment decisions

More efficiently, the Strong et al. (2014) method for calculating EVPPI (Section 2.3.2) can be extended to calculate stratum-specific and average per person EVPPI for a stratified decision [215].

Assume the value of x can be measured with certainty for every patient (i.e. the ‘test’ assigning an individual to $\mathbf{X} = x$ is accurate) and the results of the PSA for a range of values of \mathbf{X} are assumed available. Let θ be the set of all unknown parameters in the CEA model, ϕ the set of parameters of interest, and $\bar{\phi}$ the remaining parameters in θ . The EVPPI for ϕ and $\mathbf{X} = x$ can be expressed as:

$$EVPPI_x(\boldsymbol{\phi}) = E_{\boldsymbol{\phi}} \left[\max_j E_{\bar{\boldsymbol{\phi}}|\boldsymbol{\phi}} [NB(j, \boldsymbol{\phi}, \bar{\boldsymbol{\phi}}, x)] \right] - \max_j E_{\boldsymbol{\theta}} [NB(j, \boldsymbol{\theta}, x)]$$

Assume EVPPI is estimated using K PSA samples. Across the K samples x can take different values. The PSA samples, indexed by k , can be expressed as the sum of the conditional expectation and a mean-zero error:

$$NB(j, \boldsymbol{\theta}^{(k)}, x^{(k)}) = E_{\bar{\boldsymbol{\phi}}|\boldsymbol{\phi}^{(k)}} [NB(j, \boldsymbol{\phi}^{(k)}, x^{(k)})] + \varepsilon^{(k)} \quad (5.7)$$

This is the same as the method in Section 2.3.2, but x is considered as an additional regressor. Therefore, the first term in Equation 5.7 can be thought of as a function of $(\boldsymbol{\phi}^{(k)}, x^{(k)})$, say $f(j, \boldsymbol{\phi}^{(k)}, x^{(k)})$:

$$NB(j, \boldsymbol{\theta}^{(k)}, x^{(k)}) = f(j, \boldsymbol{\phi}^{(k)}, x^{(k)}) + \varepsilon^{(k)} \quad (5.8)$$

with no form imposed on $f(\cdot)$. As in Strong et al. (2014), the NMB outputs from the CEA can be seen as ‘noisy’ data to learn about $f(j, \boldsymbol{\phi}^{(k)}, x^{(k)})$. For $k = 1, \dots, K$, $NB(j, \boldsymbol{\theta}^{(k)}, x^{(k)})$ and the values for $\boldsymbol{\phi}^{(k)}$ and $x^{(k)}$ form a sample of outcome and predictor variables. Therefore, the problem can be expressed as a non-parametric regression, estimating $f(\cdot)$. Using a GAM, as outlined in Strong et al. (2014) (Section 2.3.2), a set of K fitted values $\hat{f}(\cdot)$ can be extracted to give an estimate of $E_{\boldsymbol{\theta}} [NMB(j, \boldsymbol{\theta}^{(k)}, x^{(k)})]$ for all k .

The fitted values can be partitioned into subsets for each potential x . Let K_x be the number of simulations in stratum x , then using these subsets, and the methodology presented in Section 5.3.1:

$$\widehat{EVPPI}(\boldsymbol{\phi}, x) = \frac{1}{K_x} \sum_{k=1}^{K_x} \max_j \hat{f}(j, \boldsymbol{\phi}^{(k)}, x^{(k)}) - \max_j \frac{1}{K_x} \sum_{k=1}^{K_x} \hat{f}(j, \boldsymbol{\theta}^{(k)}, x^{(k)})$$

Using w_x , the overall EVPPI, averaged over the heterogeneous population is:

$$\widehat{EVPPI}^{(X)}(\boldsymbol{\phi}) = \sum_x w_x \widehat{EVPPI}(\boldsymbol{\phi}, x)$$

To calculate the standard error for this EVPPI estimate the method in Section 2.5 can be used, which involves re-sampling the coefficients of the GAM model in Equation 5.8.

Application to data

To compare the two calculation methods presented, the EVPPI has been calculated using the case study CEA from Chapter 4 under the assumption that the distribution of the ESS (and BMI) are constant over their range.

The EVPPIs have been calculated for a stratified decision where the optimal treatment decision is decided for each strata separately. Two alternative methods are used:

- the multiple calculation method
- the single calculation method

The following sets of parameters have been explored, in each case describing the expected change in ESS due to treatment and the effect of baseline ESS on this:

- treatment with MAD; $\phi = \{\rho_{MAD}, \gamma_A^{MAD}\}$
- treatment with CPAP; $\phi = \{\rho_{CPAP}, \gamma_A^{CPAP}\}$
- treatment with MAD and with CPAP; $\phi = \{\rho_{MAD}, \gamma_A^{MAD}, \rho_{CPAP}, \gamma_A^{CPAP}\}$
- treatment with CPAP compared directly with treatment with MAD:
 $\phi = \{\rho_{MAD} - \rho_{CPAP}, \gamma_A^{MAD} - \gamma_A^{CPAP}\}.$

The CEA uses the model from Chapter 4: the network meta-regression model using AD and IPD, assuming $\gamma_A = \gamma_W$, and the adherence model presented in Chapter 3.

The results are shown in Figure 5.5 and are presented as per person EVPPIs averaged over the population distributions for ESS (and BMI). The results for the single and multiple calculation methods are similar. However, the standard error corresponding to the estimate of the EVPPI calculated using the single calculation method for the same number of PSA samples is smaller as the GAM has been estimated using a larger number of PSA samples so the uncertainty around its coefficients is smaller. Figure 5.5 indicates the most value in collecting information on the change in the treatment effect due to CM, MAD and CPAP. There is little extra value in this over the value of collecting information on the comparison between MAD and CPAP directly. The value associated with collecting more information on the impact of treatment with MAD is greater than that of CPAP, reflecting the available data.

For stratification by ESS and BMI, the results for the single and multiple parameter methods are similar, indicating most value in collecting information on the change in the treatment

effect due to MAD, compared to CPAP (Figure 5.5). Additionally, the value associated with finding more information about the impact of treatment with CPAP is greater than that of MAD reflecting the value of the existing IPD on MAD when increasing the number of stratifiers.

5.3.2 The Expected Value of Sample Information for parameters related to stratification

Having calculated the EVPPI and finding value in collecting further information on the treatment effect for both MAD and CPAP, the next logical step is to consider the EVSI for the proposed treatment comparisons. A number of different trials can be considered. Key considerations when planning a proposed study in this situation are:

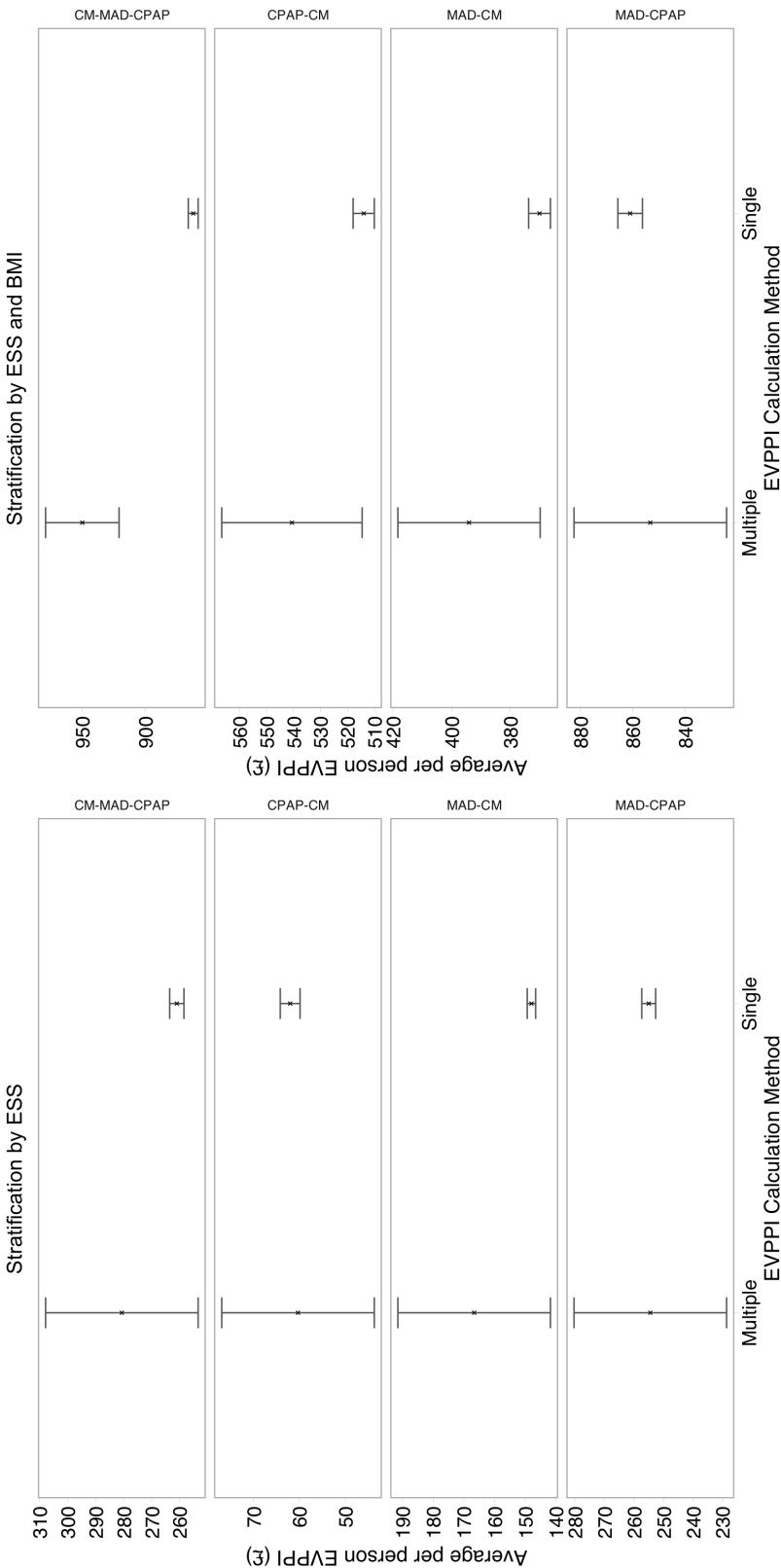
- What is the population of interest? i.e. What baseline ESS (and BMI) should the study population have?
- What treatments are being compared?
- Is the study a parallel or a cross-over trial?
- Are we planning to conduct a new study, in which case we would expect IPD to become available? Or are we proposing to search literature for additional data, which may only be available as AD?

The ESS can take values between zero and 24 [113]. From Chapter 4 the optimal treatment under current information changes at a baseline ESS of 7. There appears to be little uncertainty around the optimal treatment (CPAP) for patients with high initial ESS. Further, the literature review in Section 4.2 found many studies used populations with similar baseline ESS. It may be that further information is more valuable for specific subgroups.

Secondly, EVPPI calculations found value in collecting information comparing CM and MAD, CM and CPAP, CPAP and MAD, and a three-armed study comparing CM, MAD and CPAP. A three-armed study was optimal in terms of EVPPI, although the EVPPI for comparing CPAP and MAD directly is similar. When considering realistic study sizes, there may not be value in any treatment comparisons. EVSI calculations for all treatment comparisons with significant EVPPI should be carried out.

Both parallel and cross-over trials are regularly carried out to compare interventions for OSAHS. This has implications for study size and the results. A cross-over trial, where each

Figure 5.5 Point and 95% estimates of the Expected Value of Perfect Partial Information comparing the value of collecting further information on the treatment effect for different treatment comparisons for a population with uniform distribution(s) for the stratifier(s) of interest compared between a single and multiple calculation method



of the M participants receive all interventions where the order of allocation of interventions assumed to be randomised provides M data points for the impact of each treatment. A parallel trial with the same number of participants will, assuming equal allocation between arms, provide $\frac{M}{a}$ data points for each treatment (where a is the number of arms in the study). Additionally, while not considered here, the type of trial can impact the cost - both in terms of the monetary costs and the opportunity cost caused by delaying the new treatment regime while information is collected.

IPD is considered the ‘gold standard’ for inclusion in meta-regressions. Any future study would hopefully provide IPD to update the meta-regression. However, there may be situations where only AD would be generated, for example if a more comprehensive/systematic review of the literature was undertaken.

EVSI calculations have been carried out assuming the inclusion criteria of the proposed study defined as five ESS populations each with a ESS between:

- 0 and 10 (mild daytime sleepiness)
- 10 and 16 (moderate daytime sleepiness)
- 16 and 24 (severe daytime sleepiness)
- 0 and 24 (the whole range of ESS)
- 4 and 8 (the range of ESS where there is the most uncertainty around the optimal decision)

and that the distribution of the ESS in the resulting study population is uniform over each range.

When considering stratification by baseline ESS and BMI, three different inclusion criteria for BMI have been considered taking into account the range of BMI in the population under consideration (Chapter 4) and those at the lower and higher end of this distribution:

- BMI between $28\text{-}36\text{kgm}^{-2}$
- BMI between $28\text{-}32\text{kgm}^{-2}$
- BMI between $32\text{-}26\text{kgm}^{-2}$

The inclusion criteria for BMI and ESS have been combined, giving a total of fifteen different trial populations with the same set of studies considered for both single and multiple parameter stratification. For each trial population and treatment comparison four trial designs are considered:

1. parallel study generating IPD
2. cross-over study generating IPD
3. parallel study generating AD
4. cross-over study generating AD

Methodology by Strong et al. (2015) (Section 2.4.2) has been extended to allow for these calculations, in a similar way to EVPPI (Section 5.3.1) [216].

A multiple calculation method for the Expected Value of Sample Information relating to stratified treatment decisions

As when calculating EVPPI, the EVSI can be calculated for each stratum and aggregated to gain the average EVSI over the population of interest.

Let $\widehat{EVSI}_x(\mathbf{Y})$ be the estimate of EVSI calculated for the collection of data \mathbf{Y} , using the methods of Strong et al. (2015) (Section 2.4.2) for a stratifier x , and w_x defined as earlier [216]. The estimated average per person EVSI over the population, $\widehat{EVSI}(\mathbf{Y})$, can be calculated by:

$$\widehat{EVSI}(\mathbf{Y}) = \sum_x \widehat{EVSI}_x(\mathbf{Y}) w_x$$

The standard error, s , for the estimate of $EVSI(\mathbf{Y})$ can be calculated by:

$$\sqrt{\sum_x s_x^2 w_x^2}$$

As with the EVPPI method (Section 5.3.1) $EVSI(\mathbf{Y})$ and its standard error can be estimated by calculating $\widehat{EVSI}_x(\mathbf{Y})$ and s_x^2 for each x separately, using the methods by Strong et al. (2015) (Section 2.4.2) and aggregating [216].

Single calculation method for the Expected Value of Sample Information under a stratified treatment decision

More efficiently, the calculation method of Strong et al. (2015) can be extended to calculate the EVSI for a stratified population using a single non-parametric regression [216].

As in the extension for EVPPI (Section 5.3.1) we want to find the EVSI for a fixed x (i.e. the ‘test’ assigning an individual to a strata is accurate). Let \mathbf{Y} be the (uncollected) data with a realisation of y . Then EVSI can be expressed as:

$$EVSI(\mathbf{Y}) = E_{\mathbf{Y}} \left[\max_j E_{\boldsymbol{\theta}|\mathbf{Y}} [NB(j, \boldsymbol{\theta}, x)] \right] - \max_j E_{\boldsymbol{\theta}} [NB(j, \boldsymbol{\theta}, x)] \quad (5.9)$$

Assume K PSA samples, where x takes different values. Using Strong et al. (2015) (Section 2.4.2) the PSA samples can be expressed as the sum of the conditional expectation and a mean-zero error [216]:

$$NB(j, \boldsymbol{\theta}^{(k)}, x^{(k)}) = E_{\boldsymbol{\theta}|\mathbf{Y}} [NB(j, \boldsymbol{\theta}^{(k)}, x^{(k)})] + \varepsilon^{(k)} \quad (5.10)$$

The first term on the right hand side of Equation 5.10 takes a different value of $\boldsymbol{\theta}^{(k)}$ and $x^{(k)}$ and so can be thought of as a function of $T(\mathbf{Y})$, say $g(j, T(\mathbf{Y}^{(k)}), x^{(k)})$, where no form is imposed on $g(\cdot)$ and $T(\mathbf{Y})$ is the sufficient statistic which contains all information gained from the proposed study on the parameters of interest (Section 2.4.2) [216]. Then:

$$NB(j, \boldsymbol{\theta}^{(k)}, x^{(k)}) = g(j, T(\mathbf{Y})^{(k)}, x^{(k)}) + \varepsilon^{(k)}$$

The NMB outputs from the CEA can be seen as ‘noisy’ data to help learn about $g(\cdot)$. For all $k = 1, \dots, K$, the values of $NB(j, \boldsymbol{\theta}^{(k)}, x^{(k)})$ are known and so can be presented as a non-parametric regression problem of the NMB outputs from the PSA on the sufficient statistic $T(\mathbf{Y})$ and x . Undertaking a regression using a GAM (Section 2.4.2) the set of K fitted values $\hat{g}(\cdot)$ can be extracted [216].

These fitted values can be split into subsets based on x . Letting K_x be the number of PSA samples with stratifier x :

$$\widehat{EVSI}(\mathbf{Y}, x) = \frac{1}{K_x} \sum_{k=1}^{K_x} \max_j \hat{g}(j, T(\mathbf{Y})^{(k)}, x^{(k)}) - \max_j \frac{1}{K_x} \sum_k \hat{g}(j, T(\mathbf{Y})^{(k)}, x^{(k)})$$

For a stratifier distribution defined by w_x , the average EVSI over a stratified or heterogeneous population, can be estimated by:

$$\widehat{EVSI}(\mathbf{Y}) = \sum_{x \in \mathbf{X}} w_x \widehat{EVSI}(\mathbf{Y}, x) \quad (5.11)$$

Calculation of the specific study designs

All of the proposed studies can use the methods outlined above. However, they all generate different sets of data, \mathbf{Y} , and thus have different sufficient statistics, $T(\mathbf{Y})$, describing the information they provide on the parameters of interest. This section outlines how the sufficient statistic can be defined for each of the proposed studies.

Let the proposed study have M participants, with each individual $p = 1, \dots, M$ treated with intervention j being in stratum x_{jp} . Note that in a cross-over trial x_{jp} is the same for all j within a person p , whereas in a parallel trial each individual is indexed by treatment j and participant p within each j . Recall that each x_{jp} is assumed to arise from a discrete uniform distribution defined by the inclusion criteria for each study. For example, if the study was of patients with a baseline ESS of between 10 and 16, $x_{jp} \sim U(10, 16)$.

Given M and the set of x_{jp} , the information from the study can be expressed as the baseline ESS and change in ESS for the population, aggregated over the treatment arms or for each participant depending on the study design. Each participant, p , provides data on their observed change in ESS, y_{jp} during a period on intervention j , assumed to be generated as:

$$y_{jp} = ESS_{jp}(t_1) - ESS_{jp}(t_0) \quad (5.12)$$

where t_0 and t_1 are the baseline and follow-up time respectively and ESS_{jp} is the ESS for patient p treated with intervention j , so $ESS_{jp}(t_0) = x_{jp}$. As in Section 4.3 but dropping the study i subscript for clarity, the underlying treatment effect for patient p treated with intervention j is μ_{jp} :

$$y_{jp} \sim N(\mu_{jp}, \Phi^2) \quad (5.13)$$

$$\mu_{jp} = \nu + \rho_j + \theta z_{jp} + \gamma_A^j z_{jp} \quad (5.14)$$

where the study provides $z_{jp} = x_{jp} - \bar{x}$ and ν , ρ_j , θ and γ_A^k are the random effects means and coefficients from the meta-regression presented in Section 4.3.2. The random effects means have been used to represent the mean outcome in the new study rather than predictions from the random effects distributions. The new study is designed to have a population with similar

characteristics (for a given baseline ESS) to the case study CEA, which was broadly typical of the studies included in the meta-regression.

How this underlying data is observed depends on the study design as different study designs (such as parallel, crossover or cluster-randomised studies) can collect different data which can lead to different sufficient statistics and so different values of EVSI for the same n . The four proposed designs are presented below:

Study design 1: A parallel study generating IPD

Assume patients are equally allocated between treatment arms (a), so each treatment has $\frac{M}{a}$ participants (rounded to the nearest integer). A trial with this study design will provide data, for each patient, of the form:

1. a vector of the M baseline ESS (and BMI values if two-parameter stratification), $\{z_{jp}\}$ for $p = 1, \dots, M$.
2. y_{jp} for each patient, with $\frac{M}{a}$ patients for each j .

Data for each p is generated as above for each intervention j separately. The data produced from the study on the impact of the baseline ESS on the change in ESS can be expressed in a more concise form through fitting a simple linear regression of the form in Equation 5.14 to the study data. The change in ESS due to treatment $j = \{1, 2\}$ relative to $j = 0$ and the impact of baseline ESS on this can be written as:

$$E[y_{jp}] - E[y_{0p}] = \rho_j + \gamma_A^j z_{jp}$$

The estimated values of the coefficients ρ_j and γ_A^j , from fitting the regression to the study data contain all the information from the proposed new study comparing two interventions in its most concise form. Therefore, $T(\mathbf{Y}) = (\hat{\rho}_j, \hat{\gamma}_A^j)$.

Study design 2: A cross-over study generating IPD

This study would provide data of the following form:

1. a vector of the M baseline ESS (and BMI values if two parameter stratification), $\{x_{jp}\}$ for $p = 1, \dots, M$
2. y_{jp} for each participant and each treatment comparison generated using Equation 5.12

If we assume, as with cross-over trials in the meta-regression (Chapter 4) that data from cross-over trials is equivalent to data from a parallel trial with $2 \times M$ participants in a two-armed trial and $3 \times M$ participants in a three-armed trial due to the presence of a washout period between interventions, then the method is as in Case 1.

Study design 3: A parallel study generating AD

Study design 1 provided data for each participant p . However, here the data is aggregated over all participants. So the information provided is:

1. the average baseline ESS (and BMI if we are stratifying on two variables) across the population of M individuals in the studies of which there are M_j participants each of the A treatment arms:

$$\bar{z} = \frac{1}{A} \sum_j \frac{1}{M_j} \sum_{p=1}^M y_{jp} \quad (5.15)$$

2. as in the meta-regression (Chapter 4), when AD is produced the y_{jp} are latent and unobserved with:

$$y_j = \frac{1}{M_j} \sum_{p=1}^{M_j} y_{jp}$$

Therefore, the information from the study about the expected change in ESS due to treatment $j = \{1, 2\}$ relative to $j = 0$ is simply:

$$T(\mathbf{Y}) = y_j - y_0 = \hat{\beta}_j \quad (5.16)$$

This study alone cannot provide information on γ_A^j or ρ_j separately.

Study design 4: A cross-over study generating AD

In study design 2, IPD was provided for a cross-over trial. However, here the values need to be aggregated across participants in a similar way to in study design 3. The resulting data from the study is:

1. the average baseline ESS (and BMI if considering two-parameter stratification) across the population, \bar{z} , Equation 5.15
2. the average relative change in ESS across the population with treatment j relative to treatment $j = 0$, $\hat{\beta}_j$, Equation 5.16.

Assume, as in study design 2, that due to a washout period between interventions, we can treat the data from the cross-over trial the same as if it came from a parallel trial with $2 \times M$ arms or $3 \times M$ arms for a three-armed trial. Under this assumption the same procedure as in the case of study design 3 can then be followed.

Results of the Expected Value of Sample Information calculations

The results of the EVSI calculations for single-variable stratification are presented in Figure 5.6. There appears to be greater value for the same number of study participants for collecting data from a cross-over trial assuming it is feasible for the 2 (or 3) time periods to be considered as independent designs. Additionally, there appears to be more value in collecting information for studies where the participants have a baseline ESS of between 4 and 6; or between 0 and 10 - representing the areas where there is the most decision uncertainty. At realistic study sizes, information on trials comparing MAD and CPAP and the three armed trials appears to be most valuable with little difference between the two and three armed studies. This indicates that once study costs have been taken into consideration a two arm trial comparing MAD and CPAP would likely be optimal. It is likely costs for a cross-over trial will be greater than for a parallel trial, so while the information from a cross-over trial may be more valuable, the associated ENBS may be greater for a parallel trial.

The results for the EVSI calculations for two variable stratification are presented in Figures 5.7 and 5.8 all assuming cross-over trials are carried out. (study designs two and four only). The results for parallel trials are expected to be similar. The values for the EVSI considering a three armed trial collecting IPD are not presented. This model requires a GAM with eight parameters along with the interactions, and is therefore not well defined and difficult to calculate (Section 2.3.2). The computational burden of calculating the EVSI for two variable stratification was much greater than for one variable stratification. For this reason the multiple calculation method has been implemented in this situation, as the GAM requires four parameters and their interactions as opposed to the single calculation approach which requires six parameters and their interactions in the GAM.

As with single parameter stratification EVSI, there appears to be more value in collecting information for the ESS distributions between four and six for the same number of participants. Additionally, there appears to be more value for carrying out a study with a population with higher BMI values. Based on the original individual participant data (Figure 4.4) there is a lack of data for these higher BMI values and so we would expect the EVSI to be greater for this population. In comparison to the single parameter stratification results,

Figure 5.6 Results of the EVSI calculation for four study designs to collect further information to inform the four treatment comparisons compared between five baseline ESS distributions

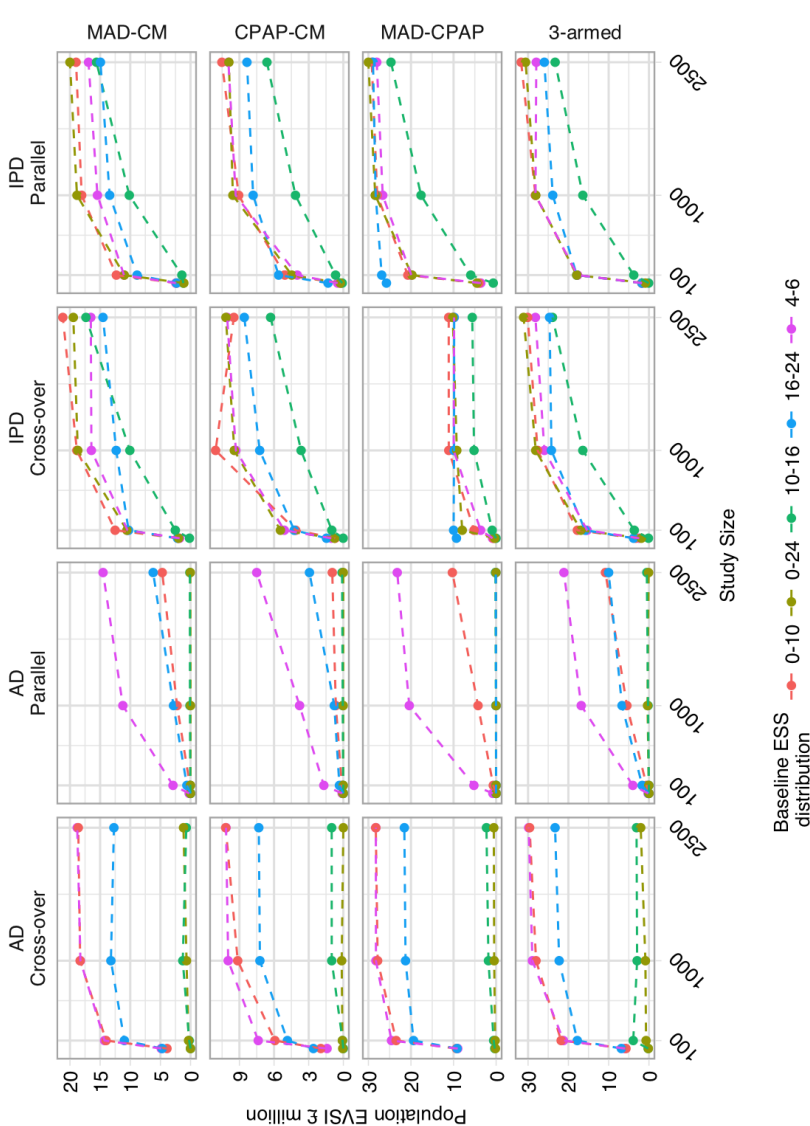


Figure 5.7 Results of the EVSI calculation for a proposed cross-over study collecting aggregated data, four treatment comparisons, five baseline ESS distributions and three BMI distributions to collect further information to inform the relationship of the treatment effect to baseline ESS and BMI

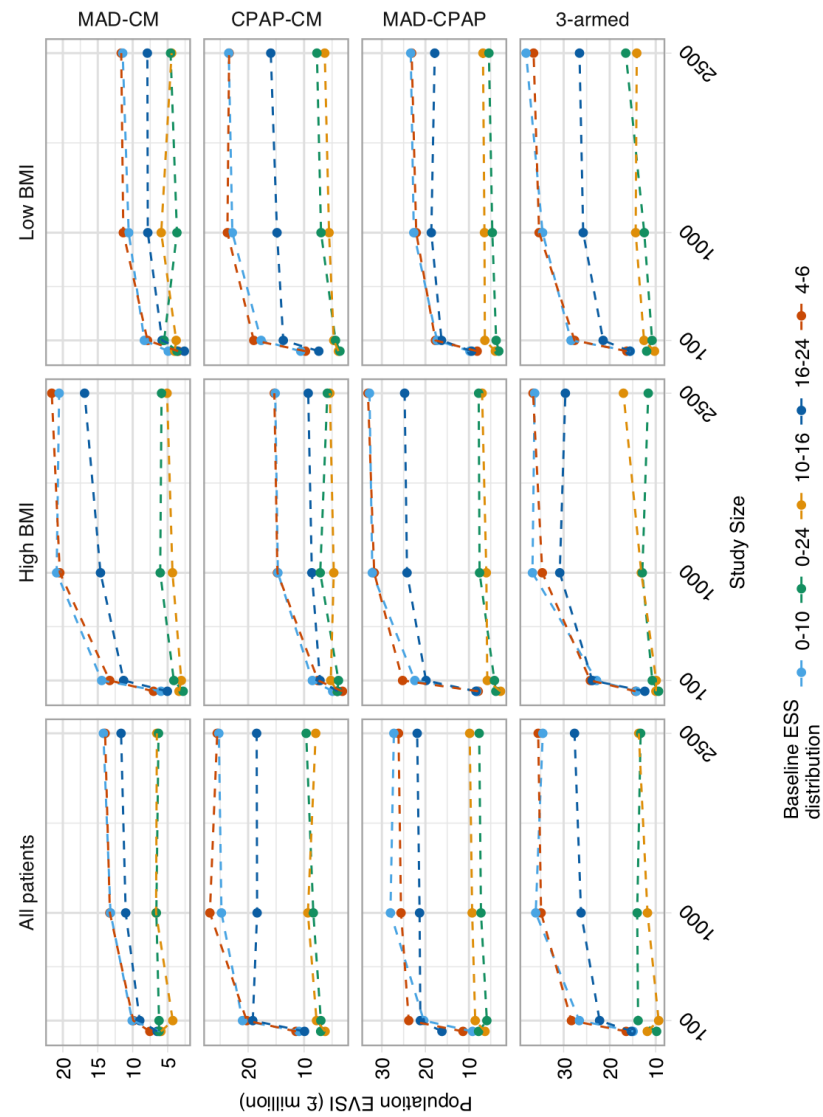
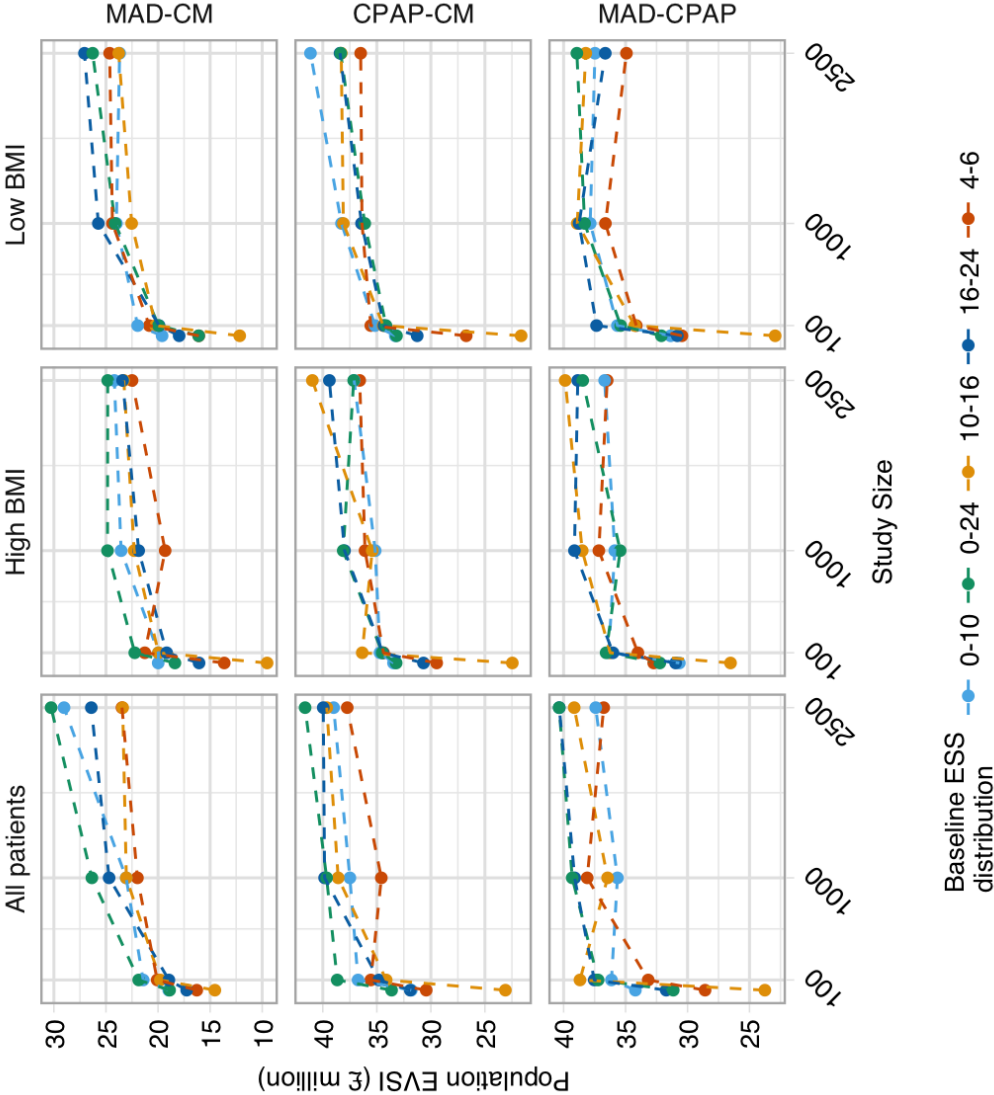


Figure 5.8 Results of the EVSI calculation for a proposed cross-over collecting individual participant data, four treatment comparisons, five baseline ESS distributions and three BMI distributions to collect further information to inform the relationship of the treatment effect to baseline ESS and BMI



there is more value in carrying out further study on the treatment comparison of CPAP and CM compared to MAD and CM. With two stratifiers IPD becomes more important in the estimation of the effect of stratification and no IPD was available for treatment with CPAP.

As explained in Chapter 4, the EVSI alone is not sufficient to make a decision on which study to undertake in practice. The study costs would also need to be considered. In terms of whether a study should stratify by one or two parameters, it is important to consider whether we believe BMI has an impact on the treatment effect. As mentioned earlier, BMI is not directly related to the treatment effect but it can influence the rate of CVEs and disease progression. Costs of the trials for one and two covariates are unlikely to be very different - there would be little extra cost in collecting someone's BMI in addition to collecting their ESS.

5.4 Discussion

This chapter builds upon Chapter 4. That found for the case study CEA some groups of the population would benefit from treatment with an intervention different to the intervention which is optimal for the population on average. This chapter explored the health economic value of implementing a stratified regime. This was presented by a more formal analysis of the costs and benefits to the population of the stratified regime. In addition, methods for calculating the value of collecting further information to guide a stratified regime have been presented and illustrated. This work could be further developed to create an optimal portfolio of interventions for patients with OSAHS subject to budget constraints and patient profiles.

Espinoza et al. (2014) developed methods to assess the 'value of heterogeneity' for subgroups [67]. These have been applied to non-categorical stratification variables - which can be continuous (such as weight) or discrete (such as ESS). The value of stratification has been calculated for the case study which shows some benefit in stratification (static value of heterogeneity) and in the value of resolving uncertainty in the parameters that explain the differences between strata (dynamic value of heterogeneity).

The value of stratification can be highly dependent on the distribution of the strata in the population: if the proportion of the population whose optimal intervention is changed due to stratification is low the value of stratification is lower. The extent to which physicians obey the stratification regime (leakage) is also an important consideration in whether to stratify

treatment.

Value of information quantities under stratification have been calculated through extensions to the Strong et al. (2014, 2015) methods [215, 216]. The single non-parametric regression methods enabled an increase in precision in the EVPPI and EVSI results for the same quantities relative to a multiple calculation method. The single non-parametric regression method for estimating EVSI was applied to the value of cross-over and parallel trial designs and the collection of IPD and AD. For all treatment comparisons the EVSI calculations showed cross-over trials were more valuable with the same number of participants than parallel trials. Additionally, collecting IPD produced a higher EVSI for the same study size compared to collecting AD.

5.4.1 Data limitations

Estimates of the value of stratification are dependent on the evidence informing the relationship of the treatment effect and the stratifiers. This also depends on the network meta-regression methods used to summarise it. The limitations of this data and its methods were discussed in Section 4.6.

The focus of this chapter has been on calculating the theoretical value of stratification. Some consideration has been made to the costs of stratification. In the case study CEA the costs of implementing the stratification regimes proposed would be low. The ESS is a short, simple, patient completed questionnaire. Similarly, little or no costs are associated with calculating BMI. However, we had no data to support this supposition other than the cost of a consultant appointment (Appendix B.8). If the stratifiers were more complex to measure then additional data would be needed to assess the costs of stratification. This may be subject to uncertainty and could potentially require modelling.

There is little information on the prevalence of each strata in the population. This is a major data limitation. The potential value of improving this information is not captured by the EVPPI or EVSI. The population distribution is a pre-specified assumption when calculating value of information quantities. Conceptually, the value of information for learning about the population distribution is different to the value of information for a parameter. In this case the aim is to learn an unknown distribution. If the population distribution could be parametrised, for example as a transformed Beta with unknown shape and scale, then standard value of information methods could be used. However, it may be difficult to express the current information on the population as a prior on the parameters of such a distribution. An alternative

method to collect further information on the population distribution could be to carry out an observational study or an expert elicitation exercise with physicians. Further, data, such as those by Health Survery for England, could be used to explore the distribution of BMI in the population. This could be potentially be available by OSAHS severity through data linkage with Hospital Episode Statistics (HES) data or CPRD data. Additionally, it is simple but time consuming to calculate the value of stratification and the value of information quantities for different population distributions to present these results as a deterministic sensitivity analysis.

In the case study CEA model no relationship is assumed between the stratifiers and the risk of adverse events due to co-morbidities. While the structure of the model was devised in collaboration with physicians, the potential for such a relationship should be explored further before stratification is implemented [134]. Phelps (1997) identified that utility may differ due to strata [163]. In this work the ESS was shown to impact the baseline utility of an individual in the CEA (Chapter 4). However, the BMI was assumed not to impact on an individual patients' utility.

Another area where data is limited is on leakage and take-up rate of the interventions. Leakage, the extent to which physicians do not adhere to the guidance, can impact the decision on whether stratification should be implemented. Coyle et al. (2003) highlighted the difficulties regarding this [47]. Explicitly including leakage in the calculations may encourage physicians to give patients close to a treatment switching threshold a choice of intervention. This is not necessarily a bad outcome. The treatment switching threshold may vary between people due to unmodelled characteristics. Taking patient preferences into consideration when they are in a stratum close to a treatment switching threshold may improve adherence to an intervention. If a patient chooses the intervention they believe they will prefer and adhere to then this may be their optimal treatment. In this case study, both treatments are currently available in the UK and so leakage is a real concern.

Additionally, Sculpher (2008) raises the concern that some stratification measures could be inappropriately influenced to gain entry into a preferred subgroup and so receive a preferred treatment [193]. As the ESS is a patient reported questionnaire it would be relatively simple for a patient to 'adjust' their answers to get a preferred treatment. Stratification on a variable such as a BMI or a blood test result would be more difficult for a patient to influence.

Similarly, the take-up rate, the rate at which hospitals and physicians implement the new treatment regime, can impact on the value of stratification. Further information on the extent to which the physicians would actually implement a stratified treatment regime would be useful. This could be collected qualitatively or quantitatively through an expert elicitation exercise with physicians. Finally, ethical constraints need to be considered when choosing stratifiers. The NICE do not permit stratification on socio-demographic characteristics such as age, sex, gender, and race as this may be considered to be inequitable [151].

5.4.2 Methodological issues

The value of stratification and the value of further research to guide stratification are all dependent on the models used to relate the stratifiers to the treatment effect. Therefore, all the methodological issues highlighted in Chapter 4 are also of concern in this chapter. In particular, the EVSI values assume the relationship between the treatment effect, μ , and the stratifiers \mathbf{X} is linear. The EVPPI and EVSI represent parameter uncertainty. Uncertainty around the linear relationship would not be identified from the value of information calculations. The value of information for structural uncertainty due to the linear relationship between the treatment effect and stratifiers could be calculated. To do this, the meta-regression model would need to be extended to include potential non-linear terms, then the value of information for these additional parameters calculated. However, due to a lack of information on the form of these non-linear terms the results may be sensitive to the priors given to these additional terms.

Population-level values of collecting further information are presented. These are a clearer representation of whether the proposed study should be implemented (Section 2.2.2) compared to individual-level values. However, scaling the EVPPI and EVSI to the population-level values comes with its own issues (Chapter 2). Additionally, population-level values require assumptions on the future trends of the distribution of the stratifiers in the population. If an increase in the incidence and/or diagnosis rate is predicted, this may increase the proportion of the population with less severe disease which could potentially change the decision regarding stratification. However, quantifying this uncertainty formally is difficult without the benefit of foresight!

In the case study CEA the expected value of collecting information on different strata were compared. Additionally, as with the EVSI for adherence to interventions (Chapter 3) the costs of carrying out the study have not been calculated. These costs can be monetary or in terms of health foregone. They may depend on the choice of the intervention, the study design, the length of study, or the number of trial participants. No allowance has been made

for how easy it would be to find patients with the required characteristics. This is another area where knowing the true population distribution of \mathbf{X} would be valuable. The lower the probability of $\mathbf{X} = x$, the more difficult or time consuming it would be to recruit patients to a study where this strata is an inclusion criteria.

The opportunity costs of delaying implementation of a treatment strategy to carry out further research have been ignored. A number of papers discuss the irreversibility of treatment decisions [13, 64, 71]. While NICE can change their treatment recommendations in light of new evidence, this may be costly to the hospital or treatment provider. This is especially true if there are large sunk costs in implementing a strategy. These costs would not be reimbursed should a treatment decision be reversed. In this application, it is unlikely to be an issue - both treatments are currently widely available either on the NHS or from dentists. Additionally, no large investments would be required to change the treatment strategy or to implement a stratified strategy. Therefore, it may be reasonable in this case to implement a stratified treatment regime on the basis of current information while collecting further evidence.

The approach of Strong et al. (2014, 2015) has been used in the EVPPI and EVSI calculations [215, 216]. Both use GAM non-parametric regression. For a large number of parameters and/or a number of interaction terms the GAM is impractical (Section 2.3.2). Strong et al. (2014) state for the calculation of EVPPI that a GP should be used [215]. The GP requires significant user input and so it is not as easy to implement (Section 2.3.2). Alternatively, a number of other methods could be used to calculate the EVPPI and EVSI (Chapter 2).

Further, it has been assumed the proposed study in the EVSI calculations will have a population like the average study used in the meta-regression through the use of the random effects means. However, if we do not believe this assumption then it may be preferable to use a 'new' study generated by the predictive distributions of the meta-regression which is able to reflect uncertainty about the study population (for a given baseline ESS). As in Chapter 3, the point estimates are expected to be similar under the two assumptions but the values estimated by the 'new' study would be more uncertain.

5.4.3 Future research priorities

As mentioned throughout this discussion there are obvious areas for future research. Information on population distributions and their evolution over time would help immensely. As mentioned earlier in this discussion, it may be possible to calculate the value of learning the distribution of the strata in the population. However, as noted by Phelps (1997), this

information would be costly to collect [163]. The cost of this research should be deducted from the value of stratification.

Knowing how much a study would cost would help assess the value of further research, as with modelling adherence to interventions (Chapter 3). Developing cost functions for trials is a research area in itself and is not a trivial exercise. The ENBS decides whether a study should be implemented, not the EVSI alone. The cost of the proposed study in terms of monetary and opportunity loss need to be considered. This would involve getting specific costs for proposed study designs. The opportunity loss due to further research would be more challenging to calculate [137].

5.5 Conclusion

Chapter 4 found evidence for the case study CEA that the optimal treatment differed for groups of the population depending on baseline disease severity measures. This chapter presented how the population-level health economic value of stratification for covariates can be calculated. The importance of reflecting the population distribution and leakage due to stratification has been illustrated. Finally, an extension to the Strong et al. (2014, 2015) EVPPI and EVSI methods for stratified, heterogeneous populations has been presented. From this, populations have been identified where further information would be most useful to guide stratification.

Chapter 6

Discussion and Conclusions

This final chapter provides an overview of the research findings in this thesis, its contributions to the literature, and its limitations. The chapter concludes by identifying areas of future research which could further develop the work presented in this thesis.

6.1 Overview of thesis findings

The overarching aim of the thesis was to explore methods for quantifying and targeting areas in CEAs where reducing uncertainty would be most beneficial. All methodological work has been illustrated using a case study CEA of interventions for patients with OSAHS introduced in Section 1.3 [198]. The primary focus was on two under-explored aspects of uncertainty in CEAs:

- **Patients' adherence to interventions:** Full adherence to interventions is often implicitly assumed. Alternatively, point estimates for adherence are used as in the case study CEA [198]. The aim of this section of work was to devise methods to combine all readily available data to present adherence rates which reflect current uncertainty around adherence. The impact of modelling adherence on the results of the case study CEA was presented. Methods for estimating the value of collecting further information on adherence were developed and demonstrated.
- **Heterogeneity between patients:** A population-level CEA can result in some population groups being treated with a suboptimal intervention. The first aim of this section was to present how evidence synthesis methods could be used to explore whether a treatment effect changes with respect to individual-patient baseline characteristics. The second aim was to estimate the population-level value of stratification using the

results of the evidence synthesis. Value of information methods illustrated the value of collecting further information on the potential stratifiers.

Chapter 1 introduced the key health economic concepts, value of information quantities, and evidence synthesis methods forming the basis of the methodology in this thesis. The case study of interventions for patients with OSAHS was presented.

Chapter 2 expanded on the theory of the value of information methods (Section 1.1.3) to present a number of recently developed methods designed to improve the efficiency in estimating the EVPPI and EVSI. These new methodological advances appear to improve the computational speed without compromising on the accuracy of the estimates. However, the number of applications of value of information methods remains low [20, 207]. A PSA uses Monte Carlo simulation to indicate the uncertainty around the optimal treatment decision. Increasing the number of PSA samples increases the accuracy of the CEA results. Chapter 2 suggested commonly used numbers of simulations are insufficient to present the EVPI to a suitable degree of accuracy.

A method to calculate the uncertainty due to the PSA sample size in EVPPI estimated using non-parametric regression was presented. In the case study CEA this had a small impact on the standard error, with most error in the estimate coming from the uncertainty in the coefficients of the non-parametric regression model. However, it is important to acknowledge the potential impact in general of Monte Carlo error due to the PSA sample size.

Population-level value of information is more useful compared to an individual patient value for guiding decisions on collecting further information. A population-level value can be compared directly to the costs of the research. Even modest uncertainty in quantities in the population-level calculation such as prevalence, incidence, and the time horizon can impact the population-level value considerably (Section 2.2.2). This can potentially alter the recommendation for future research. Additionally, uncertainty on quantities such as the uptake of interventions and diagnosis rates which are often implicitly assumed to be 100% can impact the collection of further information.

Modelling long-term adherence to interventions is a neglected area, with point estimates or implicit full adherence often assumed. Chapter 3 explored methods for combining all available data on adherence using Bayesian meta-analysis approaches for time-to-event data. This provided predictive estimates and their associated uncertainty for adherence to interventions over time through the parameters of a Weibull model. Using the results

from the meta-analysis in the case study CEA showed that adherence modelling induced greater uncertainty around the optimal intervention when compared to using point estimates for adherence. Applications of value of information methods (Chapter 2) showed value in collecting further information on adherence to both interventions, even for small sample sizes at a single timepoint.

Chapters 4 and 5 explored the second major topic of the thesis: baseline heterogeneity between patients in CEAs and its impact on the optimal treatment allocation. Baseline heterogeneity can lead to different treatments being optimal for groups of the population. Population-level decisions can lead to some groups receiving suboptimal interventions. Chapter 4 presented Bayesian meta-regression methods using all available data by combining AD and IPD to explore the impact of baseline characteristics on treatment effects. These methods found a relationship between baseline ESS and change in ESS with intervention. The addition of a second stratifier, BMI, had little impact. The inclusion of IPD improved the precision of the meta-regression results. Incorporating the results of the network meta-regression into the case study CEA showed on an individual patient-level the optimal treatment for patients with a baseline ESS of seven or less was a MAD and for the rest of the population CPAP.

Chapter 4 found benefit for an individual-patient in stratifying the optimal treatment decision on the basis of at least ESS. Chapter 5 took a health economic perspective, calculating the population-level value of stratification. The importance of the distribution of the stratifiers in the population was highlighted. When a small proportion of the population have their treatment decision changed under stratification it is more likely that the cost of stratification will outweigh its benefits. Leakage, the practice of prescribing patients suboptimal interventions under a stratified treatment regime, was found to impact the value of stratification, altering the decision of implementing a stratified regime in some circumstances.

Methods for calculating EVPPI and EVSI introduced in Chapter 2 were extended to estimate EVPPI and EVSI for a heterogeneous population. The 'new' method produces similar estimates with increased precision for similar computational cost. The value of proposed trials with different treatment comparisons was considered along with different study designs (cross-over or parallel trials) collecting AD or IPD. Population-level value was found in all cases and for small population sizes. No comparisons with the costs of running the trials were made. Due to the large population values, driven by the high incidence and prevalence rates for OSAHS, it is believed these trials would provide a positive ENBS.

6.2 Challenges and limitations

A lack of data was an issue throughout the thesis. However, quantifying uncertainty in the presence of sparse data was a key aim. Evidence synthesis methods have been used to estimate posterior distributions for model parameters of interest and to target future research. Data on MADs as treatment for OSAHS was sparse relative to information on CPAP for both adherence to the interventions (Chapter 3) and the relationship between the effect of the intervention and a measure of the baseline disease severity (Chapters 4-5). This is not an unusual situation. Data is more likely to be available on the intervention considered the ‘comparator’ in a CEA, as this is generally current best practice [150]. For a proposed ‘new’ intervention, data on efficacy and adherence would be expected to be limited. The availability of methods for use in ‘realistic’ situations is valuable and allows the potential value of further data collection to be assessed.

In a meta-regression, studies are required to have complete data on all covariates and treatment effects. This can lead to reporting bias if those not reporting all information are systematically different. Higgins et al. (2008) present methods for dealing with missing outcome data in meta-analyses [99]. Kunkel and Kaizar (2017) present a comparison of methods for multiple imputation of IPD meta-analyses and found that the priors used in the multiple imputation can impact on the meta-analysis results. [120].

The Bayesian network meta-regression (Chapter 4) found that including a small amount of IPD improved the precision of the results. However, it is notoriously difficult to access IPD. In this thesis authors were not approached for additional IPD because of the additional time required and the low likelihood of response. Additionally, it was felt presenting the benefit of a small amount of IPD would be valuable as it is a ‘realistic’ situation for researchers. The lack of IPD is an issue that many researchers face and led to the development of meta-analysis methods combining AD and IPD [181]. The value of including IPD in the meta-regression shows data owners the potential gains to the wider community that could occur if they make their data available.

A number of studies explored factors that may affect adherence to MAD and CPAP as interventions for OSAHS [4, 28, 32, 34, 79, 82, 115, 119, 121, 132, 135, 228, 233, 236]. These found a range of patient or initial usage characteristics which may affect adherence, of which disease severity were one. There was a lack of data on how adherence differed between patient groups. This information would have enabled adherence rates, dependent

on the baseline characteristics of the cohort, to be used in the case study CEA exploring stratified treatment decisions. For example, patients with a higher baseline ESS may be more likely to adhere to their intervention which would impact the treatment effect and thus the value of stratification.

A key challenge in modelling adherence to interventions (Chapter 3) was how to define what classified as ‘being adherent’ to an intervention. Many different definitions of adherence to MAD and CPAP were found in the literature with no formal, universal definition. This made it challenging to compare adherence rates between studies. It was decided to use the definition in Shapiro and Shapiro (2010): usage of the intervention for more than four hours a night for over 70% of nights [195]. This was used by a number of studies and appeared a sensible quantity of usage.

In modelling adherence (Chapter 3), adherence was assumed to be binary due to the nature of most of the data on adherence and difficulties in interpreting data presented in terms of average usage a night. A treatment effect does not drop to zero once usage of an intervention falls below a pre-defined, often arbitrary, level. It was assumed that once a patient is non-adherent to an intervention they were treated with CM (which had no treatment effect) and did not restart treatment. This may not reflect real practice. As an example, it may be that if a patient was non-adherent to a MAD their physician may start them on treatment with CPAP as suggested by the NICE guidance for OSAHS [149].

Current guidance recommends CPAP should be given to those with moderate/severe OSAHS. Those with mild OSAHS should only be prescribed CPAP once they have failed to tolerate MADs. This agrees with the findings from Chapter 4. However the work in this thesis does not take into account whether patients’ adherence to interventions is likely to be based on their disease severity. Further research, as outlined in section 3.5, would be required to answer this. This data could then be used to identify a series of treatment pathways based on patients disease severity and their likelihood of adherence to interventions.

In both the adherence meta-analysis (Chapter 3) and the stratification network meta-regression (Chapter 4) strong assumptions have been made about the form of the relationship we are trying to estimate. Both the assumption that time to non-adherence has a Weibull distribution and that there is a linear relationship between baseline ESS (and BMI) and the treatment effect were made using the available data. However, if these relationships were inappropriate the results from the evidence synthesis modelling may also be incorrect. Additional data on

adherence after more than ten years of treatment or on the change in ESS with treatment for populations with low or high baseline ESS values would help to test these assumptions. Further, more data would also enable more complex models, if required, to be well defined.

Although the ESS is a well-used measure in OSAHS research, there are biases and issues around its use. It asks patients to recall information on their sleepiness for the last ‘few days’. This is an ambiguous statement and could potentially be driven by the previous night. An objective measure would be preferable. The AHI, measured by polysomnography, is an objective measure of OSAHS severity. However, it has been found to be uncorrelated with the ESS. The case study CEA did not include AHI as a parameter. The primary outcome of interest for the CEA from a clinical perspective is how the intervention impacts on the patients’ daytime sleepiness. It is through reducing sleepiness that patients experience benefits from the interventions.

The computational demand of estimating value of information quantities is a key concern. The recently developed methods (Chapter 2) have improved the speed of calculations with little apparent loss in accuracy compared to the two-level Monte Carlo estimates. However, their accuracy and computational burden depends on the number of PSA simulations used. Calculating the EVSI for a number of different study designs, populations, study sizes, and treatment comparisons was still a time consuming task. While not considered in this thesis the costs of carrying out these studies would need to be estimated in practice. This would add further to the calculation burden.

The methods presented for calculating value of information quantities (Chapter 2) can only be used in their current form to assess the value of information on quantities which are expressed as parameters in the model. There may be other currently un-modelled quantities where additional information would be useful. For example, quantities describing model structure such as the inclusion of BMI (for example through alternative estimation of the risk of CVEs which use BMI as a covariate) in the CEA. Strong et al. (2014a) devised a ‘discrepancy approach’ method to measuring the value of structural uncertainty where uncertain parameters are added to the model, which represent departures from structural assumptions whose EVPPI can be calculated [213].

The decision to stratify is sensitive to the distribution of the strata in the population (Chapter 5). There is little data on the distribution of the ESS in the population and no guidance on how this could be estimated. This could be a disease-specific issue when considering the

high non-diagnosis rate for OSAHS. There are similar issues when collecting information on leakage and take-up rate for interventions. Section 6.4 presents some ideas for future research on this.

6.3 Contributions to the literature

Current NICE methods guidance advises that CEA models should be run until convergence with no formal definition of convergence provided [150]. This thesis has shown the impact of the number of PSA samples on the standard error of the EVPPI when estimated using non-parametric regression methods. In this case study this had little impact. However, it is hoped this work can contribute to the argument that more PSA samples are preferred. This would increase the accuracy of the CEA and the resulting value of information quantities and help to ensure optimal allocation of funding resources.

This thesis has presented extensions to the Strong et al. (2014, 2015) methods for estimating EVPPI and EVSI [215, 216]. These enable researchers to efficiently estimate the value of information for a heterogeneous population in a stratified CEA to an increased degree of accuracy, using the same set of PSA samples for each subgroup of the population.

Adherence to interventions is commonly overlooked. This thesis has presented Bayesian evidence synthesis methods using all data providing the number or proportion of patients adherent to MAD or CPAP. This analysis has provided estimates of the non-adherence rates through time and their associated uncertainty. Quantifying uncertainty on adherence probabilistically through the posterior distributions of the adherence model parameters can enable estimation of the value of collecting information on adherence. For the case study, modelling adherence increases the uncertainty around the optimal treatment decision compared to assuming adherence to the two interventions can be represented through point estimates from Kohler et al. (2010) [119, 198].

This thesis suggests patients with OSAHS and an ESS of less than or equal to seven should be treated with MAD and the remainder with CPAP (Chapter 4). This policy is roughly in line with the current NICE guidance on treatment for OSAHS (where those with moderate to severe OSAHS, defined as having an $AHI > 15$ apnoea-hypopnoea events per hour should be treated with CPAP) [149].

The population-level value of a policy of stratification depends on the prevalence of the strata in the population and the costs of stratification. Chapter 5 found different plausible distributions of the stratifiers in the population could alter the decision around stratification. Similarly, even modest change in quantities such as population size, disease incidence rate, or the disease diagnosis rate can have a large impact on the population-level value of information. It is hoped these examples along with further work (Section 6.4) can encourage researchers to acknowledge this uncertainty.

Both MAD and CPAP are well-established interventions for OSAHS. A number of studies have explored their effectiveness and CPAP has been available on the NHS for a number of years. MADs are available from dentists in England and on the NHS in Scotland. This may mean the dataset in this case is larger than for other disease areas. This would limit the generalisability of the adherence modelling work but, conversely, would motivate estimating the value of further information on adherence.

6.4 Future research priorities

There are several areas where further research could build on the work of this thesis and many are presented in the individual chapter discussions. Firstly, work from this thesis could be used alongside the findings from Bindels et al. (2016) and Steuten et al. (2013) to educate policy makers on the calculations, applications, and importance of estimating the value of information [20, 207]. Applications of value of information can guide the appropriate allocation of future research resources ensuring the limited funds are focussed on areas producing the most health economic gain to society.

No literature was found which compared the recently published EVSI methods (Chapter 2) to each another. Most of these studies were published very recently. All compared their work to the two-level Monte Carlo method using different example CEAs. All gained favourable results for their methodology in terms of accuracy and computational demand [2, 23, 95, 107, 108, 138, 216]. A study comparing the EVSI methods would be useful to ensure the ‘best’ method to estimate EVSI is used.

More research on including structural uncertainty in value of information calculations would be useful. For example, to assess the assumption of a linear relationship between a stratifier and the treatment effect. There is some literature on this, but further work would be beneficial especially if related to one of the value of information estimation methods (Chapter 2)

[106, 213]. The uncertainty in the model structure could be parameterised, for example using a ‘discrepancy approach’ (Strong et al. (2014a)) and then value of information methods used [213].

More specifically, the assumption of the linear relationship between the covariates and treatment effect could be tested by extending the meta-regression model to include extra terms (for example, a quadratic term) with a weakly informative prior. The results from a meta-regression including the extra terms can be used in the CEA and value of information measures for these covariates estimated.

If it is believed that BMI should be reflected in the case study CEA, whereas currently it is not, the CEA could be extended. The main impact of BMI is likely to be through the risk of CVEs. In the case study CEA the Framingham risk equations are used to estimate the risk of CVEs [7]. There are other risk equations, such as QRISK3, which estimate the same probabilities and include BMI as a covariate [100]. The case study CEA model could be updated to reflect the new risk equation and any difference in results between the two models assessed. Stratification on two covariates (BMI and ESS) could be implemented using this new model.

One area of future research relevant to a large amount of work in this thesis is in quantifying the typical costs of future research. These can be broader than monetary costs [164]. Developing some rules of thumb for costing further research would be beneficial, while also acknowledging study costs can differ in many ways. These include the choice of interventions, the length of the study, the size of the study, and the type of information collected. For example, if carrying out an incremental value of information analysis (Tuffaha et al. (2016)), stating a lower limit for feasible costs of future research in various contexts would enable a quick approach to assess whether the next quantity should be estimated [229]. However, before a study commences a full costing would be needed.

Population-level value of information measures were shown through a deterministic sensitivity analysis to be sensitive to the values of incidence, prevalence, time horizon, uptake, and non-diagnosis rates used in scaling per person values to population-level (Chapter 2). Sensitivity to these values is rarely acknowledged [86]. A probabilistic approach for estimating population-level values would be useful. This could be achieved by defining distributions (informed by data where possible) for the uncertain parameters enabling a range of population-level values to be calculated. A degree of accuracy for the population-level

value could be reported taking into account any correlations between parameters. This would also enable the researcher to estimate the value of collecting further information to resolve uncertainty on these parameters.

Limited information is available on leakage, the proportion of people receiving a suboptimal treatment (Chapter 5). Incorporating leakage into calculations may encourage physicians to give patients a suboptimal treatment leading to complications in its estimation [47]. Further research on patient preferences for interventions would be useful. This data could help to assess adherence to interventions: patients may be more likely to adhere to an intervention they ‘prefer’. The same data could also be used to explore which factors impact adherence and the prevalence of leakage. For example, if physicians were found to prescribe patients in strata close to a ‘switch point’ in interventions a choice, this could be used to help assess the prevalence of leakage.

Probabilistic methods could be implemented to quantify uncertainty around the distribution of strata in the population. This is different to uncertainty about parameters. In this case, we want to learn about an unknown distribution. To use value of information methods the distribution could be parametrised, for example by a transformed Beta with unknown shape and scale parameters. Setting priors for these distributions may be difficult.

6.5 Conclusion

This thesis has explored methods for quantifying and targeting areas of CEAs where reducing uncertainty was most beneficial. Modelling adherence to interventions was found to impact the optimal treatment decision compared to when point estimates are used for adherence. Value was found for collecting more information on adherence to interventions for patients with OSAHS. Methods have been developed to explore the value of stratifying the optimal treatment decision across a heterogeneous population on the basis of disease severity measures. At an individual patient-level these found different groups of patients should be offered different treatments. However, the population-level value of stratification was found to be dependent on the distribution of the strata in the population. Value of information methods have been extended to allow for efficient calculation for heterogeneous populations; finding that the value of collecting further information can differ for different proposed study populations of interest and study designs.

References

- [1] Department for Transport (DfT) Road Accident Data - GB, 2014. URL <http://www.data-archive.ac.uk>.
- [2] A. E. Ades, G. Lu, and K. Claxton. Expected value of sample information in medical decision modeling. *Medical Decision Making*, 24:207–227, 2004.
- [3] D. Alexander, D. Harper, G. Lowe, L. Macdonald, N. Pitts, S. Qureshi, and S. Twaddle. Management of Obstructive Sleep Apnoea/Hypopnoea Syndrome in Adults. (SIGN Guideline No 73). *Scottish Intercollegiate Guidelines Network*, (June):1–38, 2003.
- [4] C. Alves, J. M. P. C. Caminha, A. M. Silva, and D. Mendonça. Compliance to continuous positive airway pressure therapy in a group of Portuguese patients with obstructive sleep apnea syndrome. *Sleep & Breathing*, 16:555–562, 2012.
- [5] American Academy of Sleep Medicine Task Force. Sleep-Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research. *Sleep*, 22(5):667–689, 1999.
- [6] K. M. Anderson. A Nonproportional Hazards Weibull Accelerated Failure Time Regression Model. *Biometrics*, 47(1):281–288, 1991.
- [7] K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B. Kannel. Cardiovascular disease risk profiles. *American Heart Journal*, 121(1 Pt 2):293–298, 1991.
- [8] L. Andronis and P. M. Barton. Adjusting Estimates of the Expected Value of Information for Implementation: Theoretical Framework and Practical Application. *Medical Decision Making*, 36(3):296–307, 2016.
- [9] J. B. Arnold. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2', 2018. URL <https://cran.r-project.org/package=ggthemes>.
- [10] B. Auguie. gridExtra: Miscellaneous Functions for "Grid" Graphics, 2017. URL <https://cran.r-project.org/package=gridExtra>.
- [11] Australian Government and National Health and Medical Research Council. National Health and Medical Research Council. URL <https://www.nhmrc.gov.au>.
- [12] G. Baio. *Bayesian Methods in Health Economics*. CRC Press, 2013.
- [13] G. Baio and A. P. Dawid. Probabilistic sensitivity analysis in health economics. *Statistical Methods in Medical Research*, 24(6):615–634, 2015.

- [14] G. Baio and A. Heath. When simple becomes complicated: Why Excel should lose its place at the top table. *Global and Regional Health Technology Assessment*, 4(1): e3–e6, 2017.
- [15] G. Baio, A. Berardi, and A. Heath. Bayesian Cost Effectiveness Analysis., 2015.
- [16] F. Barbé, J. Durán-Cantolla, F. Capote, M. De La Peña, E. Chiner, J. F. Masa, M. Gonzalez, J. M. Marín, F. García-Río, J. D. De Atauri, J. Terán, M. Mayos, C. Monasterio, F. Del Campo, S. Gomez, M. S. De La Torre, M. Martinez, and J. M. Montserrat. Long-term effect of continuous positive airway pressure in hypertensive patients with sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 181: 718–726, 2010.
- [17] M. Barnes, D. Houston, C. J. Worsnop, A. M. Neill, I. J. Myktyyn, A. Kay, J. Trinder, N. A. Saunders, R. D. McEvoy, and R. J. Pierce. A randomized controlled trial of continuous positive airway pressure in mild obstructive sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 165:773–780, 2002.
- [18] A. Basu. Individualization at the heart of comparative effectiveness research: the time for i-CER has come. *Medical Decision Making*, 29:NP9–NP11, 2009.
- [19] A. Basu and D. Meltzer. Value of information on preference heterogeneity and individualized care. *Medical Decision Making*, 27:112–127, 2007.
- [20] J. Bindels, B. L. T. Ramaekers, I. C. Ramos, L. Mohseninejad, S. Knies, J. P. C. Grutters, M. Postma, M. Al, T. Feenstra, and M. A. Joore. Use of Value of Information in Healthcare Decision Making: Exploring Multiple Perspectives. *PharmacoEconomics*, 34(3):315–322, 2016.
- [21] J. C. Borel, R. Tamisier, S. Dias-Domingos, M. Sapene, F. Martin, B. Stach, Y. Grillet, J. F. Muir, P. Levy, F. Series, J. L. Pepin, J. L. Pepin, Y. Grillet, M. Sapene, F. Martin, B. Housset, and H. Pegliasco. Type of Mask May Impact on Continuous Positive Airway Pressure Adherence in Apneic Patients. *PLoS ONE*, 8(5):e64382, 2013.
- [22] Y. Bravo Vergel, S. Palmer, C. Asseburg, E. Fenwick, M. de Belder, K. R. Abrams, and M. J. Sculpher. Is primary angioplasty cost effective in the UK? Results of a comprehensive decision analysis. *Heart*, 93:1238–1243, 2007.
- [23] A. Brennan, S. A. Kharroubi, A. O’Hagan, and J. Chilcott. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Medical Decision Making*, 27:448–470, 2007.
- [24] A. H. Briggs, K. Claxton, and M. J. Sculpher. *Decision Modelling for Health Economic Evaluation*. Oxford University Press, 2006.
- [25] A. H. Briggs, B. Mihaylova, M. J. Sculpher, A. Hall, J. L. Wolstenholme, M. Simoons, J. Deckers, R. Ferrari, W. J. Remme, M. Bertrand, and K. Fox. Cost effectiveness of perindopril in reducing cardiovascular events in patients with stable coronary artery disease using data from the EUROPA study. *Heart*, 93:1081–1086, 2007.

- [26] S. L. Brilleman, C. Metcalfe, T. J. Peters, and W. Hollingworth. The reporting of treatment nonadherence and its associated impact on economic evaluations conducted alongside randomized trials: A systematic review. *Value in Health*, 19(1):99–108, 2016.
- [27] R. Brooks. EuroQol: the current state of play. *Health Policy*, 37(1):53–72, 1996.
- [28] A. Brostrom, A. Stromberg, J. Martensson, M. Ulander, L. Harder, and E. Svanborg. Association of Type D personality to perceived side effects and adherence in CPAP-treated patients with OSAS. *Journal of Sleep Research*, 16:439–447, 2007.
- [29] W. B. Brouwer, A. J. Culyer, N. J. A. van Exel, and F. F. Rutten. Welfarism vs. extra-welfarism. *Journal of Health Economics*, 27(2):325–338, 2008. ISSN 01676296. doi: 10.1016/j.jhealeco.2007.07.003.
- [30] J. Buchanan and S. Wordsworth. Welfarism Versus Extra-Welfarism: Can the Choice of Economic Evaluation Approach Impact on the Adoption Decisions Recommended by Economic Evaluation Studies? *PharmacoEconomics*, 33(6):571–579, 2015. ISSN 11792027. doi: 10.1007/s40273-015-0261-3. URL <http://dx.doi.org/10.1007/s40273-015-0261-3>.
- [31] F. Campos-Rodriguez, A. Grilo-Reina, J. Perez-Ronchel, M. Merino-Sanchez, M. Gonzalez-Benitez, and M. Beltran-Robles. Effect of continuous positive airway pressure on ambulatory BP in patients with sleep apnea and hypertension: A placebo-controlled trial. *Chest*, 129:1459–67, 2006.
- [32] F. Campos-Rodriguez, M. A. Martinez-Garcia, N. Reyes-Nuñez, I. Caballero-Martinez, C. V. Almeida-Gonzalez, P. Catalan-Serra, and N. Peña-Griñan. Long-term continuous positive airway pressure compliance in females with obstructive sleep apnoea. *The European Respiratory Journal*, 42:1255–62, 2013.
- [33] Canadian Agency for Drugs and Technologies in Health (CADTH). *Guidelines for the Economic Evaluation of Health Technologies*, 3rd Edition. 2006.
- [34] C. L. Chai, A. Pathinathan, and B. Smith. Continuous positive airway pressure delivery interfaces for obstructive sleep apnoea. *Cochrane Database of Systematic Reviews*, (4):CD005308, 2006.
- [35] C. Chai-Coetzer, Y.-M. Luo, N. Antic, X.-L. Zhang, B.-Y. Chen, Q.-Y. He, E. Heeley, S.-G. Huang, C. Anderson, N.-S. Zhong, and R. McEvoy. Predictors of long-term adherence to continuous positive airway pressure therapy in patients with obstructive sleep apnea and cardiovascular disease in the SAVE study. *Sleep*, 36:1929–1937, 2013.
- [36] K. Chin, T. Nakamura, K.-I. Takahashi, K. Sumi, H. Matsumoto, A. Niimi, S. Fukuhara, M. Mishima, and T. Nakamura. Falls in blood pressure in patients with obstructive sleep apnoea after long-term nasal continuous positive airway pressure treatment. *Journal of Hypertension*, 24:2091–2099, 2006.
- [37] K. Claxton. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, 18:341–364, 1999.

- [38] K. Claxton and J. Posnett. An economic approach to clinical trial design and research priority-setting. *Health Economics*, 5(6):513–23, 1996.
- [39] K. Claxton, M. J. Sculpher, and M. Drummond. Viewpoint A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet*, 360(9334):711–715, 2002.
- [40] K. Claxton, L. Ginnelly, M. J. Sculpher, Z. Philips, and S. Palmer. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. *Health Technology Assessment*, 8(31):1—103, iii, 2004.
- [41] K. Claxton, S. Martin, M. Soares, N. Rice, E. Spackman, S. Hinde, N. Devlin, P. C. Smith, and M. Sculpher. Methods for the estimation of the National Institute for Health and care excellence cost-effectiveness threshold. *Health Technology Assessment*, 2015. ISSN 20464924. doi: 10.3310/hta19140.
- [42] A. Cobos, A. J. Jovell, A. García-Altés, R. García-Closas, and L. Serra-Majem. Which statin is most efficient for the treatment of hypercholesterolemia? A cost-effectiveness analysis. *Clinical Therapeutics*, 21(11):1924–1936, 1999. ISSN 1879114X. doi: 10.1016/S0149-2918(00)86740-5.
- [43] W. G. Cochran. The Combination of Estimates from Different Experiments. *Biometrics*, 10(1):101–129, 1954.
- [44] D. Cook, V. GebSKI, and A. Keech. Subgroup analysis in clinical trials. *Medical Journal of Australia*, 180(6):289–291, 2004.
- [45] M. Corporation and S. Weston. doSNOW: Foreach Parallel Adaptor for the 'snow' Package, 2017. URL <https://cran.r-project.org/package=doSNOW>.
- [46] S. R. Coughlin, L. Mawdsley, J. A. Mugarza, J. P. H. Wilding, and P. M. A. Calverley. Cardiovascular and metabolic effects of CPAP in obese males with OSA. *European Respiratory Journal*, 29:720–727, 2007.
- [47] D. Coyle, M. J. Buxton, and B. J. O'Brien. Stratified cost-effectiveness analysis: A framework for establishing efficient limited use criteria. *Health Economics*, 12(August 2002):421–427, 2003.
- [48] S. E. Craig, M. Kohler, D. J. Nicoll, D. J. Bratton, A. Nunn, R. J. O. Davies, and J. R. Stradling. Continuous positive airway pressure improves sleepiness but not calculated vascular risk in patients with minimally symptomatic obstructive sleep apnoea: the MOSAIC randomised controlled trial. *Thorax*, pages 1090–1096, 2012.
- [49] L. Cui, H. M. James Hung, S. J. Wang, and Y. Tsong. Issues Related To Subgroup Analysis in Clinical Trials1. *Journal of Biopharmaceutical Statistics*, 12(3):347–358, 2002.
- [50] C. J. Currie, P. McEwan, J. R. Peters, T. C. Patel, and S. Dixon. The Routine Collation of Health Outcomes Data from Hospital Treated Subjects in the Health Outcomes Data Repository (HODaR): Descriptive Analysis from the First 20,000 Subjects. *Value in Health*, 8:581–590, 2005.

- [51] T. P. Debray, K. G. Moons, G. van Valkenhoef, O. Efthimiou, N. Hummel, R. H. H. Groenwold, and J. B. Reitsma. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Research Synthesis Methods*, 6(4):293–309, 2015.
- [52] T. P. Debray, E. Schuit, O. Efthimiou, J. B. Reitsma, J. P. A. Ioannidis, G. Salanti, and K. G. Moons. An overview of methods for network meta-analysis using individual participant data: when do benefits arise? *Statistical Methods in Medical Research*, pages 1–14, 2016.
- [53] M. S. Dennis, J. P. Burn, P. A. Sandercock, J. M. Bamford, D. T. Wade, and C. P. Warlow. Long-term survival after first-ever stroke: the Oxfordshire Community Stroke Project. *Stroke*, 24:796–800, 1993.
- [54] Department of Health. NHS Reference Costs (2011/12), 2012. URL <https://www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-2012>.
- [55] S. Dias, A. J. Sutton, N. J. Welton, C. Hall, W. Road, D. S. Unit, and R. Court. Nice DSU Technical Support Document 3 : Heterogeneity : Subgroups , Meta-Regression , Bias and Bias-Adjustment. (September):1–75, 2011.
- [56] S. Dias, N. J. Welton, A. J. Sutton, and A. E. Ades. NICE DSU Technical Support Document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. (0):1–98, 2011.
- [57] S. Dias, N. J. Welton, A. J. Sutton, and A. E. Ades. Evidence Synthesis for Decision Making 1: Introduction. *Medical Decision Making*, 33:597–606, 2013. ISSN 1552-681X. doi: 10.1177/0272989X13487604. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3704205&tool=pmcentrez&rendertype=abstract>.
- [58] H. Diener, L. Cunha, C. Forbes, J. Sivenius, P. Smets, and A. Lowenthal. European Stroke Prevention Study 2: Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. *Journal of the Neurological Sciences*, 143:1–13, 1996.
- [59] A. V. Diez-Roux. Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis. *American Journal of Public Health*, 88(2), 1998.
- [60] Division of Sleep Medicine at Harvard Medical School. Sleep Apnoea - Understanding the Results, 2011. URL <http://healthysleep.med.harvard.edu/sleep-apnea/diagnosing-osa/understanding-results>.
- [61] P. Dolan. Modeling Valuations for EuroQol Health States. *Medical Care*, 35(11): 1095–1108, 1997.
- [62] S. Donegan, P. Williamson, U. D’Alessandro, P. Garner, and C. T. Smith. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials. *Statistics in Medicine*, 32(6):914–930, 2013.
- [63] M. Drummond, M. J. Sculpher, G. Torrance, B. J. O’Brien, and G. L. Stoddart. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth edition, 2015.

- [64] S. Eckermann and A. R. Willan. Expected Value of Information and Decision Making in HTA. *Health Economics*, 16:195–209, 2007.
- [65] D. R. Elbourne, D. G. Altman, J. P. Higgins, F. Curtin, V. Worthington, and A. Vail. Meta-analyses involving cross-over trials : methodological issues. *International Journal of Epidemiology*, 31:140–149, 2002.
- [66] R. L. Ellen, S. C. Marshall, M. Palayew, F. J. Molnar, K. G. Wilson, and M. Man-Son-Hing. Systematic review of motor vehicle crash risk in persons with sleep apnea. *Journal of Clinical Sleep Medicine*, 2:193–200, 2006.
- [67] M. A. Espinoza, A. Manca, K. Claxton, and M. J. Sculpher. The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. *Medical Decision Making*, 34:951–64, 2014.
- [68] D. Evans. Hierarchy of evidence. *Journal of Clinical Nursing*, 12:77–84, 2003. ISSN 09621067. doi: 10.1046/j.1365-2702.2003.00662.x.
- [69] J. F. Faccenda, T. W. Mackay, N. A. Boon, and N. J. Douglas. Randomized placebo-controlled trial of continuous positive airway pressure on blood pressure in the sleep apnea-hypopnea syndrome. *American journal of respiratory and critical care medicine*, 163:344–348, 2001.
- [70] J. C. Felli and G. B. Hazen. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making*, 18(1):95–109, 1998.
- [71] E. Fenwick, K. Claxton, and M. J. Sculpher. Representing Uncertainty: The Role of Cost-Effectiveness Acceptability Curves. *Health Economics*, 10(May):779–787, 2001.
- [72] E. Fenwick, B. J. O’Brien, and A. H. Briggs. Cost-effectiveness acceptability curves - Facts, fallacies and frequently asked questions. *Health Economics*, 13(5):405–415, 2004.
- [73] E. Fenwick, K. Claxton, and M. Sculpher. The value of implementation and the value of information: Combined and uneven development. *Medical Decision Making*, 28(1): 21–32, 2008.
- [74] R. Fisher. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society*, (594-604):309–368, 1922.
- [75] W. W. Flemons and M. A. Reimer. Development of a disease-specific health-related quality of life questionnaire for sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 158(2):494–503, 1998.
- [76] L. M. Friedman, C. D. Furberg, and D. L. DeMets. *Fundamentals of Clinical Trials*. Springer International Publishing, 4th edition, 2010.
- [77] T. Furukawa, M. Suzuki, M. Ochiai, H. Kawashima, N. Yokoyama, and T. Isshiki. Long-term adherence to nasal continuous positive airway pressure therapy by hypertensive patients with preexisting sleep apnea. *Journal of Cardiology*, 63(4):281–285, 2014.

- [78] F. Gagnadoux, M. Vaillant, F. Goupil, T. Pigeanne, S. Chollet, P. Masson, M. P. Humeau, A. Bizieux-Thaminy, and N. Meslier. Influence of marital status and employment status on long-term adherence with continuous positive airway pressure in sleep apnea patients. *PLoS ONE*, 6(8):1–7, 2011.
- [79] W. Galetke, L. Puzzo, C. Priegnitz, N. Anduleit, and W. J. Randerath. Long-term therapy with continuous positive airway pressure in obstructive sleep apnea: adherence, side effects and predictors of withdrawal - a ‘real-life’ study. *Respiration*, 82:155–161, 2011.
- [80] A. M. Garber and C. E. Phelps. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics*, 16(1):1–31, 1997.
- [81] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall / CRC texts in statistical science, third edition, 2014.
- [82] A. Ghazal, S. Sorichter, I. E. Jonas, and E. C. Rose. A randomized prospective long-term study of two oral appliances for sleep apnoea treatment. *Journal of Sleep Research*, 18:321–328, 2009.
- [83] H. A. Glick, J. A. Doshi, S. S. Sonnad, and D. Polsky. *Economic Evaluation in Clinical Trials*. Oxford University Press, second edi edition, 2007. ISBN 9780198529972.
- [84] A. M. Gray, P. M. Clarke, J. L. Wolstenholme, and S. Wordsworth. *Applied Methods of Cost-effectiveness Analysis in Healthcare*. Oxford University Press, 2010.
- [85] S. B. Green and D. P. Byar. Using Observational data from registries to compare treatments: The fallacy of omnimetrics. *Statistics in Medicine*, 3:361–370, 1984.
- [86] S. Grimm, S. Dixon, and J. W. Stevens. Assessing the expected value of research: resolving uncertainty and improving implementation dynamics. *Health Economics & Decision Science (HEDS) Discussion Paper Series*, (15.03), 2015.
- [87] J.-M. Grouin, M. Coste, and J. Lewis. Subgroup analyses in randomized clinical trials: statistical and regulatory issues. *Journal of Biopharmaceutical Statistics*, 15(5): 869–82, 2005.
- [88] D. Gyrd-Hansen. Willingness to pay for a QALY: theoretical and methodological issues. *Pharmacoeconomics*, 23(5):423–432, 2005.
- [89] M. G. Hans, S. Nelson, V. G. Luks, P. Lorkovich, and S. J. Baek. Comparison of two dental devices for treatment of obstructive sleep apnea syndrome (OSAS). *American Journal of Orthodontics and Dentofacial Orthopedics*, 111:562–570, 1997.
- [90] T. J. Hastie and R. Tibshirani. Generalized additive models. *Statistical Science*, 1: 297–318, 1990.
- [91] A. Hatswell, A. Bullement, M. Paulen, and M. Stevenson. Probabilistic sensitivity in health economic models; how many simulations should we run? In *HESG, Summer 2017 Meeting, University of Aberdeen*.

- [92] A. Heath and G. Baio. An Efficient Calculation Method for the Expected Value of Sample Information: Can we do it? Yes, we can. 2017. URL <http://arxiv.org/abs/1709.02319>.
- [93] A. Heath, I. Manolopoulou, and G. Baio. Estimating the expected value of partial perfect information in health economic evaluations using integrated nested laplace approximation. *Statistics in Medicine*, 35(23):4264–4280, 2016.
- [94] A. Heath, I. Manolopoulou, and G. Baio. Efficient Monte Carlo Estimation of the Expected Value of Sample Information using Moment Matching. pages 1–29, 2016. URL <http://arxiv.org/abs/1611.01373>.
- [95] A. Heath, I. Manolopoulou, and G. Baio. A Review of Methods for Analysis of the Expected Value of Information. *Medical Decision Making*, 37(7):747–758, 2017.
- [96] J. P. Higgins and S. Green. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>*. 2011.
- [97] J. P. Higgins and A. Whitehead. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*, 15(24):2733–2749, 1996.
- [98] J. P. Higgins, S. G. Thompson, and D. J. Spiegelhalter. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society. Series A*, 172(1): 137–159, 2009.
- [99] J. P. T. Higgins, I. R. White, and A. M. Wood. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials*, 5:225–239, 2008.
- [100] J. Hippisley-Cox, C. Coupland, and P. Brindle. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *British Medical Journal*, 2099(May):j2099, 2017.
- [101] J. S. Hoch, A. H. Briggs, and A. R. Willan. Something old, something new, something borrowed, something blue: A framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Economics*, 11:415–430, 2002.
- [102] J. H. Hollandt and M. Mahlerwein. Nasal breathing and continuous positive airway pressure (CPAP) in patients with obstructive sleep apnea (OSA). *Sleep & Breathing*, 7(2):87–94, 2003.
- [103] T. Hoomans, E. A. Fenwick, S. Palmer, and K. Claxton. Value of information and value of implementation: Application of an analytic framework to inform resource allocation decisions in metastatic hormone-refractory prostate cancer. *Value in Health*, 12(2):315–324, 2009. ISSN 15244733. doi: 10.1111/j.1524-4733.2008.00431.x. URL <http://dx.doi.org/10.1111/j.1524-4733.2008.00431.x>.
- [104] D. Hughes, W. Cowell, T. Koncz, and J. Cramer. Methods for integrating medication compliance and persistence in pharmacoeconomic evaluations. *Value in Health*, 10(6): 498–509, 2007.

- [105] C. H. Jackson, N. G. Best, and S. Richardson. Improving ecological inference using individual-level data. *Statistics in Medicine*, 25(12):2136–2159, 2006.
- [106] C. H. Jackson, L. Bojke, S. G. Thompson, K. Claxton, and L. D. Sharples. A framework for addressing structural uncertainty in decision models. *Medical Decision Making*, 31(4):662–674, 2011.
- [107] H. Jalal and F. Alarid-Escudero. A Gaussian Approximation Approach for Value of Information Analysis. *Medical Decision Making*, 38(2):174–188, 2018.
- [108] H. Jalal, J. D. Goldhaber-Fiebert, and K. M. Kuntz. Computing Expected Value of Partial Sample Information from Probabilistic Sensitivity Analysis Using Linear Regression Metamodeling. *Medical Decision Making*, 35(5):584–595, 2015.
- [109] J. P. Jansen. Network meta-analysis of individual and aggregate level data. *Research Synthesis Methods*, 2012.
- [110] J. P. Jansen, B. Crawford, G. Bergman, and W. Stam. Bayesian meta-analysis of multiple treatment comparisons: An introduction to mixed treatment comparisons. *Value in Health*, 2008.
- [111] J. P. Jansen, R. Fleurence, B. Devine, R. Itzler, A. Barrett, N. Hawkins, K. Lee, C. Boersma, L. Annemans, and J. C. Cappelleri. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: Report of the ISPOR task force on indirect treatment comparisons good research practices: Part 1. *Value in Health*, 14(4):417–428, 2011.
- [112] S. Jauhar, M. F. Lyons, S. W. Banham, D. A. Cameron, and R. Orchardson. Ten-year follow-up of mandibular advancement devices for the management of snoring and sleep apnea. *Journal of Prosthetic Dentistry*, 99:314–321, 2008.
- [113] M. Johns. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, 14(6):540–545, 1991.
- [114] M. Johns and B. Hocking. Excessive Daytime Sleepiness Daytime Sleepiness and Sleep Habits of Australian Workers. *Sleep*, 20(10):844–849, 1997.
- [115] M. K. Johnson, R. Carter, A. Nicol, R. Paton, and S. W. Banham. Long-term continuous positive airway pressure (CPAP) outcomes from a sleep service using limited sleep studies and daycase CPAP titration in the management of obstructive sleep apnoea/hypopnoea syndrome. *Chronic Respiratory Disease*, 1:83–88, 2004.
- [116] W. B. Kannel, D. Levy, and L. A. Cupples. Left Ventricular Hypertrophy and Risk of Cardiac Failure: Insights from the Framingham Study. *Journal of Cardiovascular Pharmacology*, 10(Suppl 6.):S135–S140, 1987.
- [117] D. G. Kleinbaum and M. Klein. *Survival Analysis: A Self-Learning Text*. Springer International Publishing, third edition.
- [118] J. Knaus. snowfall: Easier cluster computing (based on snow)., 2015. URL <https://cran.r-project.org/package=snowfall>.

- [119] M. Kohler, D. Smith, V. Tippet, and J. R. Stradling. Predictors of long-term compliance with continuous positive airway pressure. *Thorax*, 65:829–832, 2010.
- [120] D. Kunkel and E. E. Kaizar. A comparison of existing methods for multiple imputation in individual participant data meta-analysis. *Statistics in Medicine*, 36(June):3507–3532, 2017.
- [121] G. E. La Piana, A. Scartabellati, L. Chiesa, L. Ronchi, P. Raimondi, M. A. Carro, S. Zibetti, and S. Aiolfi. Long-term adherence to CPAP treatment in patients with obstructive sleep apnea: Importance of educational program. *Patient Preference and Adherence*, 5:555–562, 2011.
- [122] W. Lee, S. Nagubadi, M. H. Kryger, and B. Mokhlesi. Epidemiology of Obstructive Sleep Apnea: a Population-based Perspective. *Expert Review of Respiratory Medicine*, 2(3):349–364, 2008.
- [123] J. Lim, T. Lasserson, J. Fleetham, and J. J. Wright. Oral appliances for obstructive sleep apnoea (Review). *Cochrane Database of Systematic Reviews*, (3), 2009.
- [124] Y. K. Loke, J. W. L. Brown, C. S. Kwok, A. Niruban, and P. K. Myint. Association of Obstructive Sleep Apnea with risk of serious cardiovascular events: A Systematic Review and Meta-Analysis. *Circulation*, 5:720–728, 2012.
- [125] G. Lu and A. E. Ades. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, 23(20):3105–3124, 2004.
- [126] D. Lunn, C. H. Jackson, N. G. Best, A. Thomas, and D. J. Spiegelhalter. *The BUGS Book - A Practical Introduction to Bayesian Analysis*. Chapman and Hall, 2013.
- [127] D. MacKay. Introduction to Gaussian Processes. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166, 1998.
- [128] J. J. Madan, A. E. Ades, M. Price, K. Maitland, J. Jemutai, P. Revill, and N. J. Welton. Strategies for efficient computation of the expected value of partial perfect information. *Medical Decision Making*, 34:327–42, 2014.
- [129] S. S. Mahmood, D. Levy, R. S. Vasan, and T. J. Wang. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *The Lancet*, 383(9921):999–1008, 2014.
- [130] W. M. Makeham. On the Law of Mortality and the Construction of Annuity Tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8(6):301–310, 1860.
- [131] J. M. Marin, S. J. Carrizo, E. Vincente, and A. G. N. Augusti. Long-term cardiovascular outcomes in med with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet*, 365, 2005.
- [132] M. Marklund. Predictors of long-term orthodontic side effects from mandibular advancement devices in patients with snoring and obstructive sleep apnea. *American Journal of Orthodontics and Dentofacial Orthopedics*, 129:214–221, 2006.

- [133] M. Mason, E. J. Welsh, and I. E. Smith. Drug therapy for obstructive sleep apnoea in adults. *Cochrane Database of Systematic Reviews*, (5), 2013.
- [134] C. Mc Daid, S. C. Griffin, H. L. A. Weatherly, K. Duree, M. van der Burgt, S. van Hout, J. Akers, R. J. O. Davies, M. J. Sculpher, and M. Westwood. Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis. *Health Technology Assessment*, 13(4):iii–iv, xi–xiv, 1–119, 143–274, 2009.
- [135] N. McArdle, G. Devereux, H. Heidarnajad, H. M. Engleman, T. Mackay, and N. J. Douglas. Long-term Use of CPAP Therapy for Sleep Apnea/Hypopnea Syndrome. *Americal Journal of Respiratory and Critical Care Medicine*, 159:1108–1114, 1999.
- [136] A. D. McGown, H. K. Makker, J. M. Battagel, P. R. L’Estrange, H. R. Grant, and S. G. Spiro. Long-term use of mandibular advancement splints for snoring and obstructive sleep apnoea: A questionnaire survey. *European Respiratory Journal*, 17:462–466, 2001.
- [137] C. McKenna and K. Claxton. Addressing adoption and research design decisions simultaneously: The role of value of sample information analysis. *Medical Decision Making*, 31(6):853–865, 2011.
- [138] N. Menzies. An Efficient Estimator for the Expected Value of Sample Information. *Medical Decision Making*, (April):1–13, 2015.
- [139] {Microsoft} and S. Weston. foreach: Provides Foreach Looping Construct for R, 2017. URL <https://cran.r-project.org/package=foreach>.
- [140] H. Morgenstern. Uses of Ecologic Analysis in Epidemiologic Research. *American Journal of Public Health*, 72(12):1336–1344, 1982.
- [141] S. Morris, N. Devlin, D. Parkin, and A. Spencer. *Economic Analysis in Healthcare*. Wiley, 2007.
- [142] C. A. Moyer, S. S. Sonnad, S. L. Garetz, J. I. Helman, and R. D. Chervin. Quality of life in obstructive sleep apnea: a systematic review of the literature. *Sleep Medicine*, 2:477–491, 2001.
- [143] A. T. Mulgrew, G. Nasvadi, A. Butt, R. Cheema, N. Fox, J. A. Fleetham, C. F. Ryan, P. Cooper, and N. T. Ayas. Risk and severity of motor vehicle crashes in patients with obstructive sleep apnoea/hypopnoea. *Thorax*, 2008.
- [144] M. Murad, N. Asi, M. Alsawas, and F. Alahdab. EBM - New Evidence Pyramid. *Evidence Based Medicine*, 21(0):125–126, 2016. doi: 10.1136/ebmed-2016-110401.
- [145] National Health Service. NHS Core Principles. URL <https://www.nhs.uk/NHSEngland/thenhs/about/Pages/nhscoreprinciples.aspx>.
- [146] National Institute for Health and Care Excellence. Guide to the processes of technology appraisal: Process and methods. *NICE: PMG18*, 2018. URL <https://www.nice.org.uk/process/pmg19/resources/guide-to-the-processes-of-technology-appraisal-pdf-72286663351237>.

- [147] National Institute of Health and Care Excellence. NICE - About. URL <https://www.nice.org.uk/about>.
- [148] National Institute of Health and Care Excellence. Soft-palate implants for obstructive sleep apnoea: Interventional procedures guidance. (November):1–6, 2007. URL <https://www.nice.org.uk/guidance/ipg241>.
- [149] National Institute of Health and Care Excellence. Continuous positive airway pressure for the treatment of obstructive sleep apnoea/hypopnoea syndrome - Technology Appraisal Guidance 139. 2008.
- [150] National Institute of Health and Care Excellence. Guide to the Methods of Technology Appraisal. 2013.
- [151] National Institute of Health and Clinical Excellence. *Social Value Judgements: Principles for the development of NICE guidance*. National Institute of Health and Clinical Excellence, second edition.
- [152] National Institute of Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. 2008.
- [153] NHS Choices. Obstructive Sleep Apnoea. URL <https://www.nhs.uk/conditions/obstructive-sleep-apnoea/treatment/>.
- [154] J. E. Oakley, A. Brennan, P. Tappenden, and J. Chilcott. Simulation sample sizes for Monte Carlo partial EVPI calculations. *Journal of Health Economics*, 29(3):468–477, 2010.
- [155] Office for National Statistics. Office of National Statistics: Updated to 2009-2011 Life Tables, . URL <http://http://webarchive.nationalarchives.gov.uk/20160107151617/http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%7B%7D3A77-274529>.
- [156] Office for National Statistics. 2013 Mid year population statistics, . URL <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%7B%7D253A77-322718>.
- [157] J. Ogden. QALYs and their role in the NICE decision-making process. *Prescriber*, April:41–43, 2017.
- [158] A. O’Hagan. Science, subjectivity and software (Comment on articles by Berger and by Goldstein). *Bayesian Analysis*, 1(3):445–450, 2006.
- [159] J. Pancer, S. Al-Faifi, M. Al-Faifi, and V. Hoffstein. Evaluation of Variable Mandibular Advancement Appliance for Treatment of Snoring and Sleep Apnea. *Chest*, 116: 1511–1518, 1999.
- [160] S. Patel, S. S. Kon, C. M. Nolan, A. K. Simonds, M. J. Morrell, W. D. D. Man, and R. E. Barker. Minimum clinically important difference of the Epworth Sleepiness Scale. In *European Respiratory Journal*, page 50:PA220, 2017.

- [161] Personal Social Services Research Unit (PSSRU). Unit costs of Health and Social Care, 2012. URL <http://www.pssru.ac.uk/project-pages/unit-costs/2012/index.php>.
- [162] Pharmaceutical Benefits Advisory Committee. *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee. Version 5.0*. 2016.
- [163] C. E. Phelps. Good technologies gone bad: How and why the cost-effectiveness of a medical intervention changes for different populations. *Medical Decision Making*, 17(1):107–117, 1997.
- [164] C. E. Phelps. Tackling Heterogeneity: The Full Costs Include Those of Research Necessary to Support the Analyses. *Medical Decision Making*, 34(November):944–947, 2014.
- [165] Z. Philips, K. Claxton, and S. Palmer. The half-life of truth: what are appropriate time horizons for research decisions? *Medical Decision Making*, 28(June):287–299, 2008.
- [166] D. M. Phillippo, A. E. Ades, S. Dias, S. Palmer, K. R. Abrams, and N. J. Welton. NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions To Nice Report By the Decision Support Unit. (December):1–82, 2016.
- [167] D. M. Phillippo, A. E. Ades, S. Dias, S. Palmer, K. R. Abrams, and N. J. Welton. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Medical Decision Making*, 38(2):200–211, 2018.
- [168] T. Pieters, P. Collard, G. Aubert, M. Dury, P. Delguste, and D. O. Rodenstein. Acceptance and long-term compliance with nCPAP in patients with obstructive sleep apnoea syndrome. *European Respiratory Journal*, 9:939–944, 1996.
- [169] M. Plummer. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 2003.
- [170] M. Plummer. Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3):523–539, 2008.
- [171] J. Pratt, H. Raiffa, and R. Schlaifer. *Statistical Decision Theory*. MIT Press, Cambridge, MA, 1995.
- [172] M. J. Price, N. J. Welton, A. H. Briggs, and A. Ades. Model averaging in the presence of structural uncertainty about treatment effects: Influence on treatment decision and expected value of information. *Value in Health*, 14(2):205–218, 2011.
- [173] N. M. Punjabi. The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society*, 5:136–43, 2008.
- [174] T. Quinnell, A. Clutterbuck-James, M. Bennett, J. Jordan, R. Chadwick, M. Davies, N. Oscroft, I. Smith, C. East, M. Pittman, M. Cameron, M. Morrell, M. Glover, J. Fox-Rushby, and L. Sharples. Randomised controlled trial of mandibular advancement devices for obstructive sleep apnoea (TOMADO): One year follow-up. In *Journal of Sleep Research*, page 116, 2014.

- [175] T. G. Quinnell, M. S. Bennett, J. Jordan, A. L. Clutterbuck-James, M. G. Davies, I. E. Smith, N. Oscroft, M. A. Pittman, M. Cameron, R. Chadwick, M. J. Morrell, M. J. Glover, J. A. Fox-Rushby, and L. D. Sharples. A crossover randomised controlled trial of oral mandibular advancement devices for obstructive sleep apnoea-hypopnoea (TOMADO). *Thorax*, 69(10):938–45, oct 2014.
- [176] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016.
- [177] H. Raiffa and R. Schlaifer. *Probability and Statistics for Business Decisions*. McGraw Hill, New York, 1959.
- [178] P. Ravva, M. O. Karlsson, and J. L. French. A linearization approach for the model-based analysis of combined aggregate and individual patient data. *Statistics in Medicine*, 33(9):1460–1476, 2014.
- [179] J. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, 2nd edition, 1995.
- [180] G. S. Richardson, M. A. Carskadon, W. Flagg, J. V. A. N. Den, W. C. W. Dement, and M. M. Mitler. Excessive Daytime Sleepiness in Man: Multiple Sleep Latency Measurement in narcoleptic and control subjects. *Electroencephalography and Clinical Neurophysiology*, 45(5):621–627, 1978.
- [181] R. D. Riley, M. C. Simmonds, and M. P. Look. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology*, 60(5), 2007.
- [182] R. D. Riley, P. Lambert, J. Staessen, J. Wang, F. Gueyffier, L. Thijs, and F. Bouët. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine*, 27:1870–1893, 2008.
- [183] G. V. Robinson, D. M. Smith, B. A. Langford, R. J. O. Davies, and J. R. Stradling. Continuous positive airway pressure does not reduce blood pressure in nonsleepy hypertensive OSA patients. *The European Respiratory Journal*, 27(6):1229–1235, 2006.
- [184] A. Rosengren, L. Wilhelmsen, M. Hagman, and H. Wedel. Natural history of myocardial infarction and angina pectoris in a general population sample of middle-aged men: a 16-year follow-up of the Primary Prevention Study, Göteborg, Sweden. *Journal of Internal Medicine*, 244:495–505, 1998.
- [185] D. B. Rubin. The Calculation of Posterior Distributions by Data Augmentation: Comment. *Journal of the American Statistical Association*, 82(398):543–546, 1987.
- [186] D. L. Sackett. Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents. *CHESTLEG*, 95(2):2S–4S, 1989. ISSN 0012-3692; 0012-3692. doi: 10.1378/chest.95.2. URL <http://dx.doi.org/10.1378/chest.95.2{ }Supplement.2S>.

- [187] M. Sadatsafavi, N. Bansback, Z. Zafari, M. Najafzadeh, and C. Marra. Need for speed: an efficient algorithm for calculation of single-parameter expected value of partial perfect information. *Value in Health*, 16(2):438–48, 2013.
- [188] G. Salanti, J. P. Higgins, A. E. Ades, and J. P. A. Ioannidis. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research*, 17(3):279–301, 2008.
- [189] B. Salepci, B. Caglayan, N. Kiral, E. Torun Parmaksiz, S. Sener Comert, G. Sarac, A. Fidan, and G. Aktin Gungor. CPAP adherence of patients with obstructive sleep apnea. *Respiratory Care*, 58:1467–1473, 2013.
- [190] P. Saramago, A. J. Sutton, N. J. Cooper, and A. Manca. Mixed treatment comparisons using aggregate and individual participant level data. *Statistics in Medicine*, 31(28):3516–3536, 2012.
- [191] C. Schilling, D. Petrie, M. M. Dowsey, P. F. Choong, and P. M. Clarke. The Impact of Regression to the Mean on Economic Evaluation in Quasi-Experimental Pre-Post Studies: The Example of Total Knee Replacement Using Data from the Osteoarthritis Initiative. *Health Economics*, 26(12):e35–e51, 2017.
- [192] C. H. Schmid, P. C. Stark, J. A. Berlin, P. Landais, and J. Lau. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *Journal of Clinical Epidemiology*, 2004.
- [193] M. J. Sculpher. Subgroups and heterogeneity in cost-effectiveness analysis. *PharmacoEconomics*, 26(9):799–806, 2008.
- [194] M. J. Sculpher and K. Claxton. Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty - When is there sufficient evidence? *Value in Health*, 8(4):431–446, 2005.
- [195] G. K. Shapiro and C. M. Shapiro. Factors that influence CPAP adherence: An overview. *Sleep & Breathing*, 14(August 2015):323–335, 2010.
- [196] S. K. S. Sharma, S. Agrawal, D. Damodaran, V. Sreenivas, T. Kadiravan, R. Lakshmy, P. Jagia, and A. Kumar. CPAP for the Metabolic Syndrome in Patients with Obstructive Sleep Apnea. *New England Journal of Medicine*, 365:2277–2286, 2011.
- [197] S. J. Sharp, S. G. Thompson, and D. G. Altman. The relation between treatment benefit and underlying risk in meta- analysis. *British Medical Journal*, 313:735–738, 1996.
- [198] L. D. Sharples, M. J. Glover, A. L. Clutterbuck-James, M. S. Bennett, J. Jordan, R. Chadwick, M. A. Pittman, C. East, M. Cameron, M. Davies, N. Oscroft, I. E. Smith, M. J. Morrell, J. A. Fox-Rushby, and T. G. Quinnell. Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea–hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive for o. *Health Technology Assessment*, 18(67):1–296, 2014.

- [199] L. D. Sharples, A. L. Clutterbuck-James, M. J. Glover, M. S. Bennett, R. Chadwick, M. A. Pittman, and T. G. Quinnell. Meta-analysis of randomised controlled trials of oral mandibular advancement devices and continuous positive airway pressure for obstructive sleep apnoea-hypopnoea. *Sleep Medicine Reviews*, 27:108–124, 2015.
- [200] L. D. Sharples, A. L. Clutterbuck-James, M. J. Glover, M. S. Bennett, R. Chadwick, M. A. Pittman, and T. G. Quinnell. Comments on "Meta-analysis of randomised controlled trials of oral mandibular advancement devices and continuous positive airway pressure for obstructive sleep apnoea-hypopnoea" by Sharples et al. *Sleep Medicine Reviews*, 27:39–40, 2016.
- [201] J. Shneerson, J. Wright John, and J. J. Wright. Lifestyle modification for obstructive sleep apnoea. *Cochrane Database of Systematic Reviews*, (1), 2001.
- [202] M. C. Simmonds, J. P. Higgins, L. A. Stewart, J. F. Tierney, M. J. Clarke, and S. G. Thompson. Meta-analysis of individual patient data from randomized trials: A review of methods used in practice. *Clinical Trials*, 2(3):209–217, 2005.
- [203] T. Smith, D. J. Spiegelhalter, and A. Thomas. Bayesian approaches to random effects meta analysis: a comparative study. *Statistics in Medicine*, 14(September 1993): 2685–2699, 1995.
- [204] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(4):583–616, 2002.
- [205] D. J. Spiegelhalter, K. R. Abrams, and J. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Chichester, UK, 2004.
- [206] E. Stacy. A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*, 33(3):1187–1192, 1962.
- [207] L. Steuten, G. Van De Wetering, K. Groothuis-Oudshoorn, and V. Retèl. A systematic and critical review of the evolving methods and applications of value of information in academia and practice. *PharmacoEconomics*, 2013.
- [208] L. A. Stewart and J. F. Tierney. To IPD or not to IPD? *Evaluation & the Health Professions*, 25(1):76–97, 2002.
- [209] A. A. Stinnett and J. Mullahy. Net Health Benefits : A New Framework for the Analysis of Cost-Effectiveness Analysis Uncertainty in. *Medical Decision Making*, 18 (2 Supplement):S68–80, 1998.
- [210] J. R. Stradling, J. C. T. Pepperell, and R. J. O. Davies. Sleep apnoea and hypertension: proof at last? *Thorax*, 56 Suppl 2(Suppl II):ii45–49, 2001.
- [211] M. Strong. SAVI - Sheffield Accelerated Value of Information. URL <http://savi.shef.ac.uk/SAVI/>.
- [212] M. Strong and J. E. Oakley. An Efficient Method for Computing Single-Parameter Partial Expected Value of Perfect Information. *Medical Decision Making*, 33(6): 755–766, 2012.

- [213] M. Strong and J. E. Oakley. When is a model good enough? Deriving the expected value of model improvement via specifying internal model discrepancies. *Journal of uncertainty quantification*, 2:106–125, 2014.
- [214] M. Strong, J. E. Oakley, and A. Brennan. An efficient method for computing the Expected Value of Sample Information . A non-parametric regression approach. *Medical Decision Making*, pages 1–19, 2013.
- [215] M. Strong, J. E. Oakley, and A. Brennan. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Medical Decision Making*, 34(3):311–26, 2014.
- [216] M. Strong, J. E. Oakley, A. Brennan, and P. Breeze. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample: A Fast, Nonparametric Regression-Based Method. *Medical Decision Making*, 35(July):570–583, 2015.
- [217] M. Sucena, G. Liistro, G. Aubert, D. O. Rodenstein, and T. Pieters. Continuous positive airway pressure treatment for sleep apnoea: compliance increases with time in continuing users. *The European Respiratory Journal*, 27(4):761–766, 2006.
- [218] P. W. Sullivan and V. Ghushchyan. Preference-Based EQ-5D Index Scores for Chronic Conditions in the United States. *Medical Decision Making*, 26:410–420, 2006.
- [219] S. Sundaram, J. Lim, and T. J. Lasserson. Surgery for obstructive sleep apnoea in adults (Review). *The Cochrane Collaboration*, 1(4):1–76, 2013.
- [220] A. J. Sutton and K. R. Abrams. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, 10(4):277–303, 2001.
- [221] A. J. Sutton, D. Kendrick, and C. A. C. Coupland. Meta-analysis of individual and aggregate level data. *Statistics in Medicine*, 27(May 2007):651–669, 2008.
- [222] A. Tarasiuk, S. Greenberg-Dotan, T. Simon-Tuval, A. Oksenberg, and H. Reuveni. The effect of obstructive sleep apnea on morbidity and health care utilization of middle-aged and older adults. *Journal of the American Geriatrics Society*, 56:247–254, 2008.
- [223] S. Taylor. *Health Psychology*. McGraw Hill, Toronto, 2006.
- [224] The Nordic Cochrane Centre. Review Manager (RevMan), 2014.
- [225] H. H. Thom, G. Capkun, A. Cerulli, R. M. Nixon, and L. S. Howard. Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an application to pulmonary arterial hypertension. *BMC Medical Research Methodology*, 15(1):1–16, 2015.
- [226] K. M. Thompson and J. D. Graham. Going Beyond the Single Number: Using Probabilistic Risk Assessment to Improve Risk Management. *Human and Ecological Risk Assessment*, 2(4):1008–1034, 1996.

- [227] S. G. Thompson and J. P. Higgins. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21:1559–1573, 2002.
- [228] T. Tokunaga, T. Ninomiya, Y. Kato, Y. Ito, T. Takabayashi, M. Tokuriki, H. Sunaga, and S. Fujieda. Long-term compliance with nasal continuous positive airway pressure therapy for sleep apnea syndrome in an otorhinolaryngological office. *European Archives of Oto-Rhino-Laryngology*, 270:2267–73, 2013.
- [229] H. W. Tuffaha, L. G. Gordon, and P. A. Scuffham. Value of Information Analysis Informing Adoption and Research Decisions in a Portfolio of Health Care Interventions. *Practice*, 1:1–11, 2016.
- [230] R. M. Turner, R. Z. Omar, M. Yang, H. Goldstein, and S. G. Thompson. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, 19(24):3417–3432, 2000.
- [231] A. Van Gestel, J. Grutters, J. Schouten, C. Webers, H. Beckers, M. Joore, and J. Severens. The role of the expected value of individualized care in cost-effectiveness analyses and decision making. *Value in Health*, 15(1):13–21, 2012.
- [232] B. A. Van Hout, M. J. Al, G. S. Gordon, and F. F. H. Rutten. Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics*, 3(May):309–319, 1994.
- [233] M. van Zeller, M. Severo, A. C. Santos, and M. Drummond. 5-Years APAP adherence in OSA patients – Do first impressions matter? *Respiratory Medicine*, 107(12):2046–2052, 2013.
- [234] J. P. Vezina, M. B. Blumen, I. Buchet, C. Hausser-Hauw, and F. Chabolle. Does propulsion mechanism influence the long-term side effects of oral appliances in the treatment of sleep-disordered breathing? *Chest*, 140:1184–1191, 2011.
- [235] B. Vrijens, S. De Geest, D. A. Hughes, K. Przemyslaw, J. Demonceau, T. Ruppar, F. Dobbels, E. Fargher, V. Morrison, P. Lewek, M. Matyjaszczyk, C. Mshelia, W. Clyne, J. K. Aronson, and J. Urquhart. A new taxonomy for describing and defining adherence to medications. *British Journal of Clinical Pharmacology*, 73(5):691–705, 2012.
- [236] R. E. Waldhorn, T. W. Herrick, M. C. Nguyen, a. E. O’Donnell, J. Soderro, and S. J. Potolicchio. Long-term compliance with nasal continuous positive airway pressure therapy of obstructive sleep apnea. *Chest*, 97:33–38, 1990.
- [237] J. E. Ware and C. D. Sherbourne. The MOS 36-item short form health survey (SF-36): I. Conceptual framework and Item Selection. *Medical Care*, 30(6):473–483, 1992.
- [238] T. E. Weaver, A. M. Laizner, L. K. Evans, G. Maislin, D. K. Chugh, K. Lyon, P. L. Smith, A. R. Schwartz, S. Redline, A. I. Pack, and D. Dinges. An Instrument To Measure Functional Status Outcomes for Disorders of Excessive Sleepiness. *Sleep*, 20(10), 83(September):835–843, 1997.
- [239] M. C. Weinstein, B. J. O’Brien, J. Hornberger, J. Jackson, M. Johannesson, C. McCabe, B. R. Luce, and ISPOR Task Force on Good Research Practices–Modeling Studies. Principles of good practice for decision analytic modeling in health-care evaluation:

- report of the ISPOR Task Force on Good Research Practices–Modeling Studies. *Value in Health*, 6(1):9–17, 2003.
- [240] N. J. Welton and H. H. Z. Thom. Value of Information: We’ve Got Speed, What More Do We Need? *Medical Decision Making*, pages 1–3, 2015.
- [241] N. J. Welton, A. E. Ades, D. M. Caldwell, and T. J. Peters. Research prioritization based on expected value of partial perfect information: A case-study on interventions to increase uptake of breast cancer screening. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 171(4):807–834, 2008.
- [242] N. J. Welton, J. J. Madan, D. M. Caldwell, T. J. Peters, and A. E. Ades. Expected Value of Sample Information for Multi-Arm Cluster Randomized Trials with Binary Outcomes. *Medical Decision Making*, 34(April):352–65, 2013.
- [243] N. J. Welton, M. O. Soares, S. Palmer, A. E. Ades, D. Harrison, M. Shankar-Hari, and K. M. Rowan. Accounting for Heterogeneity in Relative Treatment Effects for Use in Cost-Effectiveness Models and Value-of-Information Analyses. *Medical Decision Making*, 44, 2015.
- [244] I. White. *MVMETA: Stata module to perform multivariate random-effects meta-analysis*. Statistical Software Components S456970, Boston College Department of Economics, 2007.
- [245] H. Wickham. Reshaping Data with the {reshape} Package. *Journal of Statistical Software*, 21(12):1–20, 2007.
- [246] H. Wickham. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1):1–29, 2011.
- [247] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [248] H. Wickham, R. Francois, L. Henry, and K. Muller. *dplyr: A Grammar of Data Manipulation*, 2017. URL <https://cran.r-project.org/package=dplyr>.
- [249] A. R. Willan and S. Eckermann. Optimal Clinical Trial Design using value of information methods with imperfect implementation. *Health Economics*, 19:549–561, 2010.
- [250] A. Williams. Economics of coronary artery bypass grafting. *British Medical Journal*, 291(6495):600, 1985.
- [251] E. C. F. Wilson. A Practical Guide to Value of Information Analysis. *Pharmacoeconomics*, 33:105–121, 2014.
- [252] H. Woehrle, A. Graml, and G. Weinreich. Age- and gender-dependent adherence with continuous positive airway pressure therapy. *Sleep Medicine*, 12(10):1034–1036, 2011.
- [253] S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(1):3–36, 2011.

-
- [254] World Health Organisation. Constitution of the World Health Organisation. *Basic Documents, Forty-fifth edition, Supplement*, 2006.
- [255] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr. The occurrence of sleep-disordered breathing among middle-aged adults. *The New England Journal of Medicine*, (328):1230–5, 1993.
- [256] T. Young, L. Evans, L. Finn, and M. Palta. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep*, 20(9): 705–706, 1997.
- [257] T. Young, P. E. Peppard, and D. J. D. Gottlieb. Epidemiology of obstructive sleep apnea: A population health perspective. *American Journal of Respiratory and Critical Care Medicine*, 165(9):1217–1239, 2002.

Appendix A

The Epworth Sleepiness Scale

The Epworth Sleepiness Scale (ESS) is a patient reported outcome measure published by Johns in 1991 [113]. It aims to provide a measure of the patient's level of daytime sleepiness. It consists of eight questions where patients are asked how likely they would have been to have fallen asleep in a range of everyday situations over the last few days. These situations are:

- Sitting and reading
- Watching TV
- Sitting inactive in a public place (for example, in a theatre or a meeting)
- As a passenger in a car for an hour without a break
- Lying down in the afternoon when the circumstances permit
- Sitting and talking to someone
- Sitting quietly after a lunch without alcohol
- In a car when stopped for a few minutes in traffic

Patients are asked to score each of these situations from zero (would never doze) to three (high chance of dozing). The ESS is the sum of the responses to the questions - the ESS ranges from 0 to 24.

Typically patients are recorded as having [113]:

- Mild Daytime Sleepiness if $ESS \leq 11$

- Moderate Daytime Sleepiness if $11 < ESS \leq 16$
- Severe Daytime Sleepiness if $ESS > 16$

Appendix B

A cost-effectiveness model of treatments for patients with mild-moderate OSAHS

B.1 Background

As OSAHS is a chronic disease it has morbidities which can only be adequately reflected by a lifetime cost-effectiveness model [134, 198]. There is evidence OSAHS is related to hypertension meaning patients are likely to have a greater risk of CVEs [124, 256]. Additionally, EDS can lead to an increased risk of an RTA due to impaired vigilance [143]. As all of these events are relatively rare they are unlikely to be adequately represented in a short-term CEA, for example the TOMADO within trial CEA had a four week follow-up [198].

Although TOMADO only considered MADs as a potential intervention for patients with OSAHS, the NICE Technology Appraisal Number 139 defined CM¹, MADs, and CPAP as appropriate treatment options for the decision population [149, 198]. Therefore, the CEA considers all three treatment options to best assist policy makers.

McDaid et al. (2009) developed a model, hereafter ‘the McDaid model’, to investigate the lifetime cost-effectiveness of CM, MAD, and CPAP as treatments for patients with OSAHS [134]. The structure of this model, simplified in Figure B.1, was developed using a systematic review for clinical effectiveness, consultation of existing CEA literature, and clinical expert opinion. This model followed NICE methods guidance and the NICE reference case [152]².

¹CM is a one-off GP consultation offering lifestyle advice

²Since the McDaid model was developed NICE have updated their methods guidance [150, 152].

One of the aims of McDaid et al. (2009) was "*to determine the ... cost-effectiveness of CPAP devices for the treatment of OSAHS compared with best supportive care, placebo and dental devices*" [134]. This agrees with the aims of Sharples et al. (2014) and so the McDaid model was used as the basis for the CEA [134, 198]. The Sharples et al. (2014) model updates the parameters for new evidence and focuses on the population with mild-moderate OSAHS (in terms of AHI) [198].

B.2 Model structure

The model structure in the case study is the same as in the McDaid model [134]. McDaid et al. (2009) developed a state transition cohort Markov model comparing CM, MAD, and CPAP as long-term interventions for individuals with OSAHS as part of a NICE HTA [134, 149].

Figure B.1 shows a simplified, diagrammatic representation of the CEA model. All patients start in the *OSA* state and can remain there until death unless an adverse event occurs. The *pCHD* state reflects the increased mortality and morbidity from a prior, acute, non-fatal CHD event. Only the first non-fatal CHD event is modelled. Patients can remain in the *pCHD* state until death or until they have either a non-fatal stroke or a (non-)fatal RTA. If an RTA was non-fatal then patients return to their previous state. RTAs are considered instantaneous events, as opposed to states which last a full cycle. After a non-fatal stroke patients move to a post-stroke state, *pStroke*, where they can remain until death or have a non-fatal RTA. Similar to the *pCHD* state, the *pStroke* state reflects the increased mortality and morbidity from an acute non-fatal stroke. Transitions to the CHD state after a stroke are not possible. As with CHD events, second and subsequent strokes are not modelled. Additionally, a proportion of strokes are considered disabling meaning these patients are assumed to be unable to drive and so cannot have a RTA.

B.3 Inputs to the cost-effectiveness model

B.3.1 Baseline characteristics

The case study model takes a hypothetical cohort of individuals through the model which has annual cycles with baseline patient and model characteristics in Tables B.1 and B.2 [198].

Table B.1 Baseline characteristics for the cohort used in the case study cost-effectiveness analysis

Characteristic	Value	Source
Age (years)	51	TOMADO mean [198]
SBP (mmHg)	130	TOMADO mean [198]
Smoker (0=no, 1=yes)	0	Assumption based on TOMADO [198]
Total Cholesterol (mg/dl)	224	Coughlin et al. (2007) [46]
HDL Cholesterol (mg/dl)	43	Coughlin et al. (2007) [46]
Diabetic (0=no, 1=yes)	0	Assumption based on TOMADO [198]
ECG-LVH ¹ (0=no, 1=yes)	0	TOMADO assumption [198]
Baseline ESS ²	11.9	TOMADO mean [198]

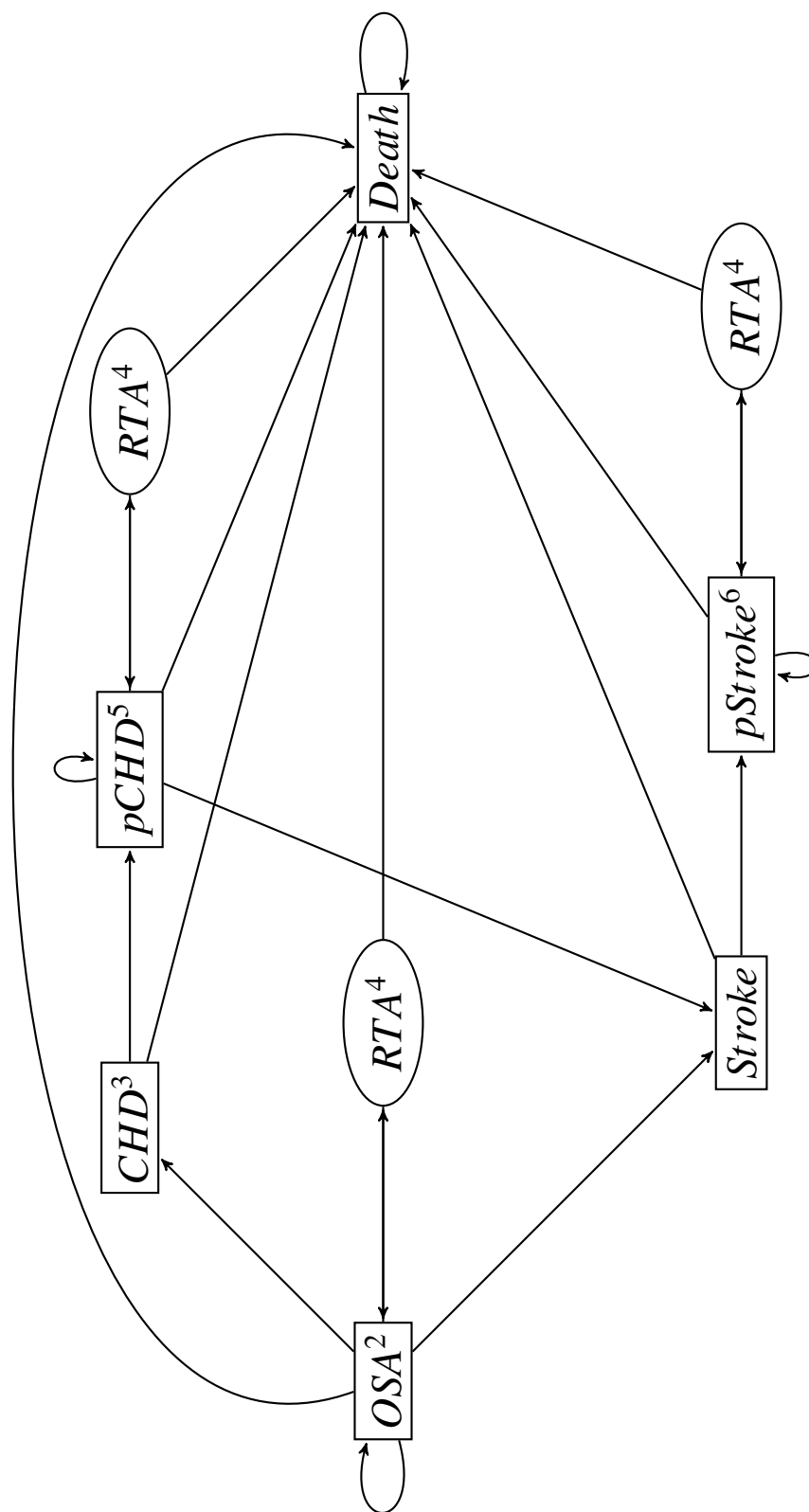
¹Left Ventricular Hypertrophy confirmed by an ECG (LVH-ECG): A condition where the wall of the left ventricle becomes thickened developed in response to high blood pressure or a heart condition [116].

²Epworth Sleepiness Scale: A measure of how sleepy an individual is during the day calculated from a questionnaire asking about the likelihood of falling asleep during a number of everyday situations giving a score from 0-24 [113] (Section 1.3.2 and Appendix A).

Table B.2 Baseline model parameters for the case study cost-effectiveness analysis

Parameter	Value	Source
Number of Cycles	65	-
Discount rate		
Costs	3.5%	NICE (2008) [152]
Utilities	3.5%	NICE (2008) [152]

Figure B.1 Simplified¹ version of the lifetime cost-effectiveness model structure developed by McDaid et al. (2009) to explore interventions for patients with OSAHS [134]



¹ Patients are allowed to become non-adherent to their intervention during the first ten years. This simplified structure ignores the presence of adherent and non-adherent states and the transitions between them. ² All patients start in the OSA state; ³ CHD event state; ⁴ Non-fatal Road Traffic Accident; ⁵ Post-CHD event state; ⁶ Post-Stroke state

B.3.2 Cardiovascular risk and sleep apnoea

CHD events and strokes, referred to as CVEs, were identified as major sources of mortality and morbidity in patients with OSAHS in McDaid et al. (2009) [134]. However, due to a lack of primary data on long-term outcomes for treatment the relationship between CVEs and OSAHS is provided by the Framingham risk score [7].

Framingham risk score equations

The Framingham risk score equations measure the risk of developing cardiovascular disease through a series of equations assuming patients are initially free of cardiovascular disease [7]. There are a number of equations each associated with a different cardiovascular endpoint. Of interest in this model are the equations using SBP³ for CHD,⁴ CHD death,⁵ Stroke,⁶ and CVD⁷.

These equations use a number of risk factors for cardiovascular events - age, gender, SBP, total cholesterol, High Density Lipoprotein (HDL) cholesterol, smoking status, diabetic status, and presence of ECG-LVH to generate hazard rates. Anderson (1991) recommend these equations are used to predict the event probabilities over a 4-12 year period [7]. The hazard ratios and probabilities were generated using a parametric statistical model based on risk factor levels and (censored) time-to-event data from the Framingham Heart Study and Framingham Offspring study cohorts (5,573 participants) [129].

Letting T be the time until the event of interest, $\mathbf{X} = \{x_1, \dots, x_k\}$ be the risk factor measurements for an individual, $\boldsymbol{\beta} = \{\beta_0, \beta_1, \dots, \beta_k\}$, and $\boldsymbol{\theta} = \{\theta_0, \theta_1\}$ be the estimated parameters⁸ for a specific endpoint [6, 7]. An accelerated failure time generalised Weibull survival model is used with the dispersion parameter depending on the location parameter [6]. That is, assume T is the time until the event of interest and the logarithm of T has location and

³There are also prediction equations using Diastolic Blood Pressure (DBP) in place of SBP

⁴including Myocardial Infarction (MI), CHD death, angina pectoris (chest pain of discomfort due to CHD and coronary insufficiency) and coronary insufficiency (a range of conditions associated with sudden reduced blood flow to the heart including MI)

⁵sudden or non-sudden

⁶including Transient Ischemic Attack (TIA) (a temporary disruption in the blood flow to the brain often called a mini-stroke)

⁷including CHD, stroke, TIA, congestive heart failure, and Peripheral Vascular Disease (PVD) (a blood circulation disorder causing the blood vessels outside of the heart and brain to narrow, block or spasm.)

⁸Within Anderson (1991) no estimates of uncertainty around the estimates of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, were presented, just point estimates were given [7].

dispersion parameters μ and σ respectively. The probability of not having an event by time t , $S(t)$ is:

$$S(t) = P(T > t) = \exp(-\exp(u(t)))$$

with:

$$\begin{aligned} u(t) &:= \frac{\log(t) - \mu}{\sigma} \\ \mu &= \beta' \mathbf{X} \\ \log(\sigma) &= \theta_0 + \theta_1 \mu \end{aligned}$$

When $\theta_1 = 0$ the model is a proportional hazards model with an underlying Weibull distribution [6]. $S(t)$ is calculated for each endpoint for patients aged from 50 to 74 with the hazard rate assumed constant over the cycle. The data in the Framingham study was only available for participants up to age 74 [7]. For ages above 74, the CEA model assumed the participants were aged 74 [134, 198]. All other covariates in the Framingham equations are set to their baseline values. Let s = stroke, c = CHD, v = CVE, and the subscripts f = fatal, nf = non-fatal. For a set of covariates \mathbf{X} and parameters $\boldsymbol{\beta}$ with $\boldsymbol{\theta}$ corresponding to the time T_Y until the endpoint of interest $Y \in \{s_{nf}, c_{nf}, v_{nf}, s_f, c_f, v_f\}$ ⁹ the probability of having the event Y in the next year is:

$$\begin{aligned} p(Y) &= P(T_Y < 1 | y, x, \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= 1 - \exp(-\exp(u(1))) \end{aligned}$$

For each cycle x is constant (apart from age) for all endpoints. $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ depend only on the endpoint. The probabilities (Table B.3) are fed into the cost-effectiveness model. The Framingham equations are important in the CEA as they govern how the treatment affects the risk of CVEs (Section B.3.8).

Other parameters relating to Cardiovascular risk

After a CVE an individual has an increased risk of mortality. The relative risk for death following CVEs are taken from Rosengren et al. (1998) and Dennis et al. (1993) [53, 184]. These were long term observational studies providing estimates of the increased risk of

⁹For clarification s_{nf} = non-fatal stroke; c_{nf} = non-fatal CHD; v_{nf} = non-fatal CVE; s_f = fatal stroke; c_f = fatal CHD; and v_f = fatal CVD

Table B.3 Quantities calculated using the Framingham equations¹ for use in the case study cost-effectiveness analysis

Quantity	Value ¹	CEA Notation ²
$p(s_f \text{ or } c_f)$	$p(c_f) + p(v_f)$	e_f
$p(s \text{ or } c)$	$p(c) + p(s)$	
$p(e_f (s \text{ or } c))$	$e_f \times p(s \text{ or } c)$	
$p(s_{nf} \text{ or } c_{nf})$	$1 - p(e_f (s \text{ or } c)) \times e_f$	e_{nf}
$p(e_f c)$	$\frac{p(c_f)}{p(c)}$	
$p(c_{nf})$	$(1 - p(c_f)) \times p(c)$	
$p(v_f)$	$p(v_f)$	s_f
$p(s_{nf})$	$(1 - p(s_f)) \times p(s)$	s_{nf}
$q_{c_{nf}}$	$\frac{p(c)}{p(c) + p(s)}$	$q_{c_{nf}}$
$q_{s_{nf}}$	$\frac{p(s)}{p(c) + p(s)}$	$q_{s_{nf}}$

¹These are expressed in terms of the quantities calculated from the Framingham equations. The β and θ values depend on the endpoint. X is constant across endpoints, within the same cycle. Further, the hazard is assumed to remain constant over the cycle.

² s_{nf} = non-fatal stroke; c_{nf} = non-fatal CHD; v_{nf} = non-fatal CVE; s_f = fatal stroke; c_f = fatal CHD; and v_f = fatal CVD

Table B.4 CHD and stroke parameters and their distributions used in the case study cost-effectiveness analysis

Parameter	Mean	SE	Distribution	Source
Relative risk of $c_f c_{nf}$	3.2	0.30	Log-normal	Rosengren et al. (1998) [184]
Relative risk of $e_f s_{nf}$	2.3	0.19	Log-normal	Dennis et al. (1993) [53]
Proportion of disabling s_{nf}	0.309	-	-	Diener et al. (1996) [58]

mortality following CVEs [53, 184]. The proportion of disabling strokes, meaning the patient was no longer able to drive and unable to have an RTA, was taken from a large RCT of over 6,000 patients which compared interventions for secondary prevention of vascular events (Diener et al. (1996)) [58, 134]. These values and their associated distributions are presented in Table B.4.

B.3.3 Risk of a road traffic accident

The values for the risk of an RTA in the general population are taken from the 2010 data from the Department of Transport [1]. This follows the methodology in McDaid et al. (2009) (Table B.5) using the numbers of UK driving licenses, fatal RTAs, and non-fatal RTAs

Table B.5 Underlying annual risk of road traffic accidents used in the case study cost-effectiveness analysis

Parameter	Value	Source
Rate of non-fatal RTAs		
Males	0.0062	Department of Transport (2014) [1]
Females	0.0053	Department of Transport (2014) [1]
Rate of fatal RTAs		
Males	7.11×10^{-5}	Department of Transport (2014) [1]
Females	2.91×10^{-5}	Department of Transport (2014) [1]

causing minor or serious injuries [134]. The reduction in the risk of RTAs for patients with OSAHS treated with MADs or CPAP are presented in Section B.3.8.

B.3.4 Utilities

Using methodology to similar McDaid et al. (2009) the IPD from the TOMADO was used to estimate a relationship between ESS and utility (based on the EQ-5D-3L) using a linear mixed effects regression model [27, 134]. The resulting algorithm maps the change in ESS with treatment to a change in utility (Table B.6) [198]. A one unit increase in ESS was found to reduce utility by 0.0061. The estimated covariance matrix for this regression was used for sampling in the PSA.

The baseline utility (at age 50) used the mean baseline ESS of 11.9 for the TOMADO population [198]. The decrements for age, stroke, and CHD events were taken from Sullivan and Gushchyan (2006) [218]. The utility after a RTA was based on EQ-5D-3L data from a repository recording information on patients' utility six weeks after an inpatient episode resulting from a RTA [50]. These values are presented in Table B.6 [134].

B.3.5 Adherence to interventions

Compliance, or adherence, to the intervention was taken from a study by Kohler et al. (2010) reporting long-term usage of CPAP (Figure B.2) with information available for adherence to ten years usage [119]. The case study found little clear evidence presenting adherence with MADs [198]. Therefore, adherence was assumed to be the same for both treatment with MAD and CPAP. This assumption was challenged in a sensitivity analysis. Due to a lack of data after ten years no change in adherence was assumed - i.e. if a patient was adherent

Table B.6 Utility estimates used in the case study cost-effectiveness analysis

Utility	Mean	SE	Distribution	Source
Baseline				
Constant	0.9094	0.0220	-	TOMADO regression [198]
ESS (per unit)	-0.0061	0.0020	-	TOMADO regression [198]
Decrement				
MAD ¹	$\Delta ESS \times -0.006$	-	-	TOMADO regression [198]
CPAP ¹	$\Delta ESS \times -0.006$	-	-	TOMADO regression [198]
Stroke	-0.0524	0.0002	Normal	Sullivan et al. (2006) [218]
CHD	-0.0635	0.0001	Normal	Sullivan et al. (2006) [218]
Age ²	-0.0007	-	-	Sullivan et al. (2006) [218]
RTA	0.62	0.27	Gamma	Currie et al. (2005) [50]

¹These quantities are defined in Appendix B.3.8, where ΔESS is the difference in ESS with intervention and CM.

²Per year from age 50

after ten years they would remain so. Additionally, adherence is assumed to be binary with an individual not being able to restart treatment after becoming non-adherent. Chapter 3 explains an alternative method, developed for this thesis, to model adherence to interventions.

B.3.6 Non-cardiovascular mortality rates

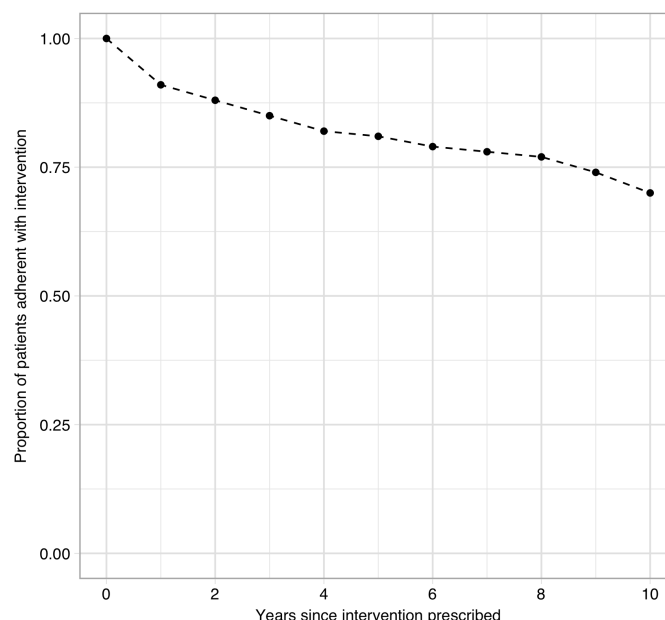
Data on the population mortality rates was taken from interim life tables (2009-2011) and mortality statistics from the Office of National Statistics [155]. The interim tables give age and gender specific all-cause mortality rates. These all-cause hazard rates were reduced according to the proportion of people who died from CHD or ischaemic heart disease. The resulting annual probabilities of death are shown in Figure B.3. The underlying mortality rates for patients suffering a CVE are calculated using the Framingham equations (Section B.3.2). As only first CVE is modelled, the increased mortality from prior events is increased by the respective relative risk (Table B.4).

B.3.7 Costs

Resource use costs

As in McDaid et al. (2009) a variety of sources were used for costs associated with the interventions (Tables B.7 - B.9) [134]. In addition, the lifespan of a CPAP machine was assumed to be seven years and a MAD was assumed to last one year. As CM was a one-off

Figure B.2 The probability individuals are adherent to their intervention¹ in the case study cost-effectiveness analysis based on data from Kohler et al. (2010) [119]



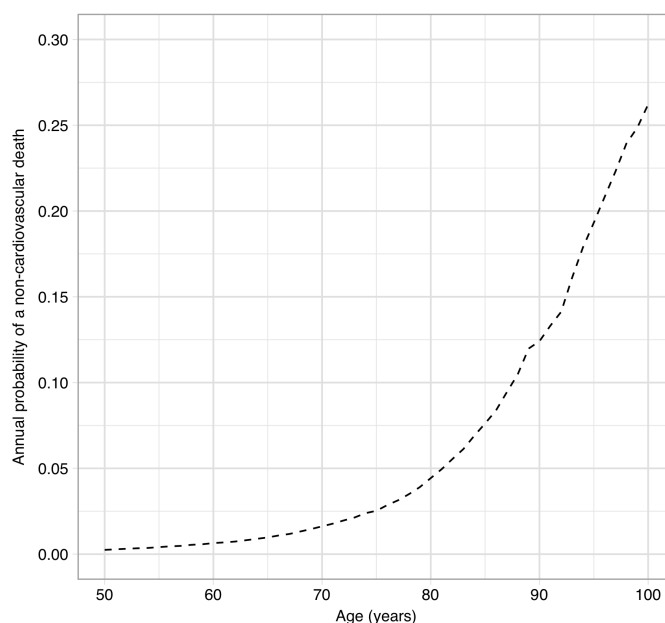
¹The lines are for reference only, it may be that adherence is not linear between the data points. However, the model uses annual cycles and so only uses the adherence rates at each point.

consultation with a GP this was assumed to be the initial cost of CM with no further costs in subsequent years. Costs were in 2011/12 prices using the Personal Social Services Research Unit (PSSRU) price indices where necessary to update the costs used in the McDaid model [134, 161].

Costs of cardiovascular events and road traffic accidents

Costs due to CVEs were taken from Briggs et al. (2007) as in McDaid et al. (2009) [25, 134]. This study used data from a large trial extrapolated using a Markov Model (n=12,218) to estimate ongoing and event costs for fatal CVEs and non-fatal CHD. The cost of stroke was taken from Bravo-Vergel et al. (2007) [22]. They used long-term data from the Nottingham Heart Attack Registry to give information on the frequency, timings, and resource use of events. Costs associated with RTAs were taken from the Department of Transport estimates of NHS costs for fatal and non-fatal RTAs in 2004 updated to 2011/12 prices [1, 161]. These are all shown in Table B.10.

Figure B.3 The annual probability of a non-cardiovascular death¹ used in the case study cost-effectiveness analysis for a male alive at age 50² using 2009-2011 interim life tables from the Office of National Statistics [155]



¹Non-cardiovascular death is defined as all cause mortality with deaths due to CHD and ischaemic heart disease removed.

²The CEA runs until the cohort is aged 115. The lifetables give data up to age 100, past this the annual probability of a non-cardiovascular death is assumed to be the same as if aged 100.

Table B.7 Costs associated with treatment by CM or MAD in 2011/2012 prices

Cost	Mean (£)	SE (£)	Distribution	Source
CM				
Initial cost	36			PSSRU (2012) [161]
MAD				
Initial cost	128			TOMADO ¹ [198]
On-going cost	105.89	47.08	Gamma	NHS Reference Costs [54]

¹Semi-bespoke device

Table B.8 Costs associated with initial use of CPAP in 2011/2012 prices

Cost	Mean (£)	SE (£)	Distribution	Source
<i>Cost of follow-up outpatient visit</i>	73.06			
Unit cost of visit	105.89	47.08	Gamma	NHS Reference costs [54]
P(follow-up appointment)	0.69	0.3	Beta	McDaid et al. (2009) [134]
<i>Cost of APAP for dose titration</i>	3.06			
p(APAP)	0.81	0.19	Beta	McDaid et al. (2009) [134]
p(home titration)	0.99	0.01	Beta	McDaid et al. (2009) [134]
Cost of APAP Machine	499			TOMADO [198]
Number of times machine used	163			McDaid et al. (2009) [134]
<i>Cost of CPAP for dose titration</i>	1.41			
p(CPAP)	0.19			McDaid et al. (2009) [134]
Cost of CPAP machine	230			TOMADO [198]
<i>Cost of in-home titration</i>	2.72			
<i>Cost of inpatient titration</i>	7.23			
P(inpatient titration)	0.01			McDaid et al. (2009) [134]
Cost of sleep study follow-up	722.8	263.5	Gamma	NHS Reference costs [54]
<i>Cost of nurse appointment</i>	44.5			
P(specialist nurse)	1			McDaid et al. (2009) [134]
Cost of appointment	44.5			PSSRU (2012) [161]
<i>Cost of titration by consultant</i>	42.37			
P(consultant for titration)	0.4	0.4	Beta	McDaid et al. (2009) [134]
Cost of consultant appointment	105.89	47.08	Gamma	NHS Reference Costs [54]
Cost of technician appointment	11.23			McDaid et al. (2009) [134]
Total initial cost	174.94			

Table B.9 Costs associated with ongoing use of CPAP in 2011/2012 prices

Cost	Mean (£)	SE (£)	Distribution	Source
<i>Annual Cost of machine</i>	36.34			$\frac{\text{Cost of CPAP}}{\text{Annuity Factor}}$
Interest Rate	3.5%			NICE [150]
Life of CPAP machine	7			McDaid et al. (2009) [134]
<i>Annual Cost of mask</i>	92.43			
Cost of mask	36.34			McDaid et al. (2009) [134]
Estimated life of mask	1			McDaid et al. (2009) [134]
<i>Annual sundries</i>	17.33			McDaid et al. (2009) [134]
<i>Annual follow-up</i>	105.89	47.08	Gamma	NHS Reference Cost [54]
<i>Total ongoing annual cost</i>	251.99			

Table B.10 Costs associated with coronary heart disease, stroke, and road traffic accidents in 2011/2012 prices

Event	Mean (£)	SE (£)	Distribution	Source
Fatal CVE	3,561	434	Normal	Briggs et al. (2007) [25]
CHD				
Year 1	11,786	505	Normal	Briggs et al. (2007) [25]
Year 2+	886	138	Normal	Briggs et al. (2007) [25]
Stroke				
Year 1	10,476	347	Normal	Bravo-Vergel et al. (2007) [22]
Year 2	2,764	334	Gamma	Bravo-Vergel et al. (2007) [22]
RTA				
Non-fatal	3,120	1,942	Gamma	Department of Transport (2004)[1]
Fatal	6,297	1,942	Gamma	Department of Transport (2004) [1]

B.3.8 Treatment effects

One treatment effect used in the cost-effectiveness model is the difference in the mean ESS over patients when treated with MAD or CPAP compared to treatment with CM:

$$\Delta ESS_{T-CM} = ESS_T - ESS_{CM}$$

where $T \in \{MAD, CPAP\}$ and ESS_T is the mean ESS score with intervention T . ESS_{CM} is the ESS for a population treated with CM. These are ‘differences in differences in ESS where:

$$ESS_T = ESS_T^{(t_1)} - ESS_T^{(t_0)}$$

where $ESS_T^{(t)}$ is the mean ESS for the population at time t with t_0 being the baseline and t_1 the follow-up time (i.e. the start and end of the period spent on treatment with T). Due to randomisation in the RCTs these baseline values are expected to cancel. ΔESS_{T-CM} is entered into the model as an uncertain parameter whose distribution is taken from the meta-analysis of RCTs in Sharples et al. (2014) for patients with mild - moderate OSAHS [198]. Chapter 4 explains how the model is extended to allow for stratified treatment effects.

A second treatment effect is the difference in mean SBP at follow-up due to treatment with MAD or CPAP compared to treatment with CM:

$$\Delta SBP_{T-CM} = SBP_T - SBP_{CM}$$

Table B.11 Modelled treatment effects and their distributions used in the case study cost-effectiveness analysis

Parameter	Mean	SE	Distribution	Source
Change in ESS				
MAD vs CM	-1.62	0.38	Normal	TOMADO HTA [198]
CPAP vs CM	-1.61	0.34	Normal	TOMADO HTA [198]
Change in SBP				
MAD vs CM	-1.13	0.53	Normal	TOMADO HTA [198]
CPAP vs CM	-2.36	0.66	Normal	TOMADO HTA [198]
Risk of RTA				
MAD vs CM ¹	0.167			McDaid et al. (2009) [134]
CPAP vs CM ¹	0.168	0.033	Log-normal	McDaid et al. (2009) [134]

¹Base-case risk, in PSA this value varies with the ratio of ESS due to treatment with MAD or CPAP

where $T \in \{MAD, CPAP\}$ and SBP_T is the mean SBP with intervention T . Similarly, SBP_{CM} is the mean SBP at follow-up for a population treated with CM. As with ΔESS_{T-CM} these are ‘difference in difference’ values. The distribution for the treatment effect was taken from the meta-analysis of RCTs in Sharples et al. (2014) [198]. The values for the whole population with OSAHS have been used due to insufficient evidence on the population with mild-moderate OSAHS.

The risk of a RTA while using CPAP has been taken from McDaid et al. (2009) [134]. The risk of a RTA while using a MAD is estimated as the risk of an RTA when using a CPAP multiplied by the ratio of the change in ESS when treated with MAD and CPAP (Table B.11):

$$RTA_{MAD-CM} = RTA_{CPAP-CM} \frac{\Delta ESS_{MAD-CM}}{\Delta ESS_{CPAP-CM}}$$

B.4 Transition probabilities

Patients move through the model (Figure B.1) for 65 annual cycles (Table B.1). The allowed transitions are presented in Table B.12 where the subscript A represents a state where members are adherent to their intervention and \bar{A} represents a state where members are not adherent to their intervention.

Patients can have a non-fatal RTA in each cycle. These are instantaneous events associated with a fixed cost and utility decrement as opposed to states. Patients return to their previous

Table B.12 Schematic table showing allowed transitions¹ in the case study cost-effectiveness analysis

		State at the end of the cycle ²									
		$OSAHS_A$	CHD_A	$pCHD_A$	$Stroke_A$	$pStroke_A$	$OSAHS_{\bar{A}}$	$CHD_{\bar{A}}$	$pCHD_{\bar{A}}$	$Stroke_{\bar{A}}$	$pStroke_{\bar{A}}$
State at the start of the cycle ²	$OSAHS_A$	✓	✓		✓		✓	✓		✓	
	CHD_A			✓	✓				✓	✓	
	$pCHD_A$			✓	✓				✓	✓	
	$Stroke_A$					✓					✓
	$pStroke_A$					✓					✓
	$OSAHS_{\bar{A}}$						✓	✓		✓	
	$CHD_{\bar{A}}$								✓	✓	
	$Stroke_{\bar{A}}$										✓
	$pStroke_{\bar{A}}$										✓
	$Dead$										✓

¹Movement between adherent and non-adherent states (A, \bar{A}) are only allowed in the first ten years; patients treated with *CM* cannot move between A and \bar{A} states.

² $OSAHS$ = OSA state; CHD = CHD state; $pCHD$ = post-CHD event state; $Stroke$ = Stroke state; $pStroke$ = post-Stroke state; $Dead$ = Dead state

state post RTA. Patients can have multiple RTAs throughout the model. Additionally, a proportion of deaths are due to fatal RTAs.

Transitions between the adherent and non-adherent states are only allowed for the first ten cycles and when the intervention is *MAD* or *CPAP*. If the intervention is *CM* then there are no transitions from A to \bar{A} states as under this model it is not possible to be non-adherent to *CM* (Section B.3.5). When a patient is non-adherent to their intervention they are assumed to be treated with *CM*. Transitions between \bar{A} states are the same as the transitions between the A states when the treatment is *CM*. No transitions are allowed from the \bar{A} states to the A states. Once an individual is non-adherent they can not become adherent again.

B.4.1 Formulation of the transition probabilities

In all cases $P(X|Y, a, T)$ represents the probability of being in state X at time i for a person in state Y at time $i - 1$ where $i = 1, \dots, 65$. The covariate T represents the intervention where $T \in \{CM, MAD, CPAP\}$. The covariate $a \in \{A, \bar{A}\}$ represents whether the individual is adherent to their intervention or not respectively. For each probability, T must be the same in cycle i and $i - 1$, but a may change from A to \bar{A} or remain the same at the next cycle

Table B.13 Notation used to derive the transition probabilities for the case study cost-effectiveness analysis

Parameter	Definition	Reference
mr	Non-CVD mortality rate	Figure B.3
e_f	Probability of a fatal CVE event	Table B.3
e_{nf}	Probability of a non-fatal CVE event	Table B.3
RR_c	Relative risk of non-stroke death following a CHD event	Table B.4
RR_s	Relative risk of all-cause death following a stroke	Table B.4
s_{nf}	Probability of a non-fatal stroke event	Table B.3
s_f	Probability of a fatal Stroke	Table B.3
s_{dis}	Probability a stroke is disabling	Table B.4
v_f	Probability of a fatal CVD event	Table B.3
r_f	Probability of a fatal RTA	Table B.5
r_{nf}	Probability of a non-fatal RTA	Table B.5
q_c	Proportion of CVE events due to CHD	Table B.3
q_s	Proportion of CVE events due to Stroke	Table B.3
A	Probability adherent with intervention	Figure B.2
\bar{A}	Probability non-adherent to the intervention ($1 - A$)	Figure B.2

($a \in \{a_0, a_1\}$ indicates the adherence state at the start and end of the cycle). All transition probabilities are cycle/age dependent, but for ease of notation the cycle notation has been omitted. The age of the cohort is used in the Framingham equations, the mortality, and the utility rates.

Transitions between A and \bar{A} states are only allowed for the first ten cycles, and when $T \in \{MAD, CPAP\}$ (Section B.3.5). In transitioning between A and \bar{A} states all patients are assumed to become non-adherent at the end of the cycle and so subject to the risks and probabilities in that cycle as though they were adherent.

Notation

The set of states, X , has elements:

$$\{OSAHS, CHD, pCHD, Stroke, pStroke, Dead\} = \{O, C, pC, S, pS, D\}$$

\bar{D} is defined to include all ‘alive’ states. Being in state X not adherent to intervention T is equivalent to being in state X and treated with CM. The notation used in the transition probabilities is shown in Table B.13.

In addition, the probabilities of non-cardiovascular death can be defined as:

$$\begin{aligned}\bar{v}_f &= p(\bar{v}_f) = 1 - e^{-mr} \\ \bar{v}_f|c_{nf} &= p(\bar{v}_f|c_{nf}) = 1 - e^{-mr \times RR_c} \\ \bar{v}_f|s_{nf} &= p(\bar{v}_f|s_{nf}) = 1 - e^{-mr \times RR_s}\end{aligned}$$

where $p(\bar{v}_f|x_f)$ is the probability of non-cardiovascular death given a prior non-fatal stroke or CHD event. These probabilities are independent of the intervention, T .

Let the probability of not dying due to non-cardiovascular causes independent of treatment, over the cycle be:

$$\begin{aligned}\bar{d} &= p(\bar{d}) = 1 - p(\bar{v}_f) \\ \bar{d}|c_{nf} &= p(\bar{d}|c_{nf}) = 1 - p(\bar{v}_f|c_{nf}) \\ \bar{d}|s_{nf} &= p(\bar{d}|s_{nf}) = 1 - p(\bar{v}_f|s_{nf})\end{aligned}$$

The model depends on the probability of becoming non-adherent in the next cycle for the first ten cycles (Section B.3.5). If an individual becomes non-adherent in the current cycle then $T = CM$ going forward.

Probability of death

$$\begin{aligned}p(D|O, a, T) &= \bar{v}_f + v_{fa}^{(T)} + r_{fa}^{(T)} \\ \left. \begin{aligned}p(D|C, a, T) \\ p(D|pC, a, T)\end{aligned} \right\} &= \bar{v}_f|c_{nf} + v_{fa}^{(T)} + r_{fa}^{(T)} \\ \left. \begin{aligned}p(D|S, a, T) \\ p(D|pS, a, T)\end{aligned} \right\} &= \bar{v}_f|s_{nf} + (1 - s_{dis}) \times r_{fa}^{(T)} \\ p(D|D, a, T) &= 1\end{aligned}$$

Probability of remaining in OSA state, i.e. no adverse event occurs

$$p(O|O, a, T) = \bar{d} \times \left(1 - e_{nfa_0}^{(T)}\right)$$

Probability of a non-fatal CHD Event

The CEA model only allows a CHD event from the OSA state. Second and subsequent CHD events are not modelled.

$$P(C|O, a, T) = \bar{d} \times e_{nf} \times q_c$$

Probability of being in the post-CHD state

$$\left. \begin{array}{l} P(pC|C, a, T) \\ P(pC|pC, a, T) \end{array} \right\} = \bar{d} c_{nf} \times [1 - e_{nf}^{(T)} \times q_s]$$

Probability of a non-fatal stroke

The CEA model only allows a stroke from the OSA or CHD state. Second and subsequent strokes are not modelled.

$$\begin{aligned} p(S|O, a, T) &= \bar{d}^{(T)} \times e_{nf}^{(T)} \times q_s \\ \left. \begin{array}{l} p(S|C, a, T) \\ p(S|pC, a, T) \end{array} \right\} &= \bar{d} c_{nf}^{(T)} \times e_{nf}^{(T)} \times q_s \end{aligned}$$

Probability of being in the post-Stroke state

$$\left. \begin{array}{l} p(pS|S, a, T) \\ p(pS|pS, a, T) \end{array} \right\} = \bar{d} s_{nf}$$

Probability of a road traffic accident

RTAs are not a state in the transition matrix, they are an instantaneous event associated with a certain cost and utility decrement, i.e. each individual is in a state and could also have a RTA event in each cycle. The probability of a non-fatal RTA in cycle i depends on the state at the end of cycle $i - 1$. Letting $Y \in \{f, nf\}$ then:

$$\left. \begin{array}{l} p(RTA_Y|O, a, T) \\ p(RTA_Y|C, a, T) \\ p(RTA_Y|pC, a, T) \end{array} \right\} = r_Y^{(T)}$$

$$\left. \begin{array}{l} p(RTA_Y|S, a, T) \\ p(RTA_Y|pS, a, T) \end{array} \right\} = [1 - s_{dis}] \times r_Y^{(T)}$$

Calculating the proportion of the population in each state at the end of a cycle

Once the transition matrix has been defined for cycle i , the proportion of the population in each state at the end of cycle $i - 1$ can be multiplied by the transition matrix to give the proportion of the population in each state at the end of cycle i . Let P be the transition matrix with elements $p_{xy} = p(y|x)$. Let q_i be the vector of the expected proportion of the population in each state at time i . Then:

$$q_{i+1} = q_i^T P$$

B.4.2 Utilities

Each state is assigned a utility. In addition, a utility is attached to a non-fatal RTA. The derivation of the utility for each state is given in Table B.6. Let $q_i(X^{(T,a)})$ be the proportion of the population in state X with treatment T and adherence state a at the end of cycle i .

For each state, there is a common contribution to the utility of:

$$U_i^{(T)} = bU + dU^{(T,a)} + dU_{age}(i - 1)$$

where bU is the baseline utility, $dU^{(T,a)}$ is the utility decrement due to treatment and dU_{age} is a utility decrement due to age (Table B.6). Then for each state:

$$U(O|a, T) = U^{(T)}$$

$$\left. \begin{array}{l} U(C|a, T) \\ U(pC|a, T) \end{array} \right\} = U^{(T)} + U_C$$

$$\left. \begin{array}{l} U(S|a, T) \\ U(pS|a, T) \end{array} \right\} = U^{(T)} + U_S$$

where U_C is the utility decrement due to a CHD and U_C is the utility decrements due to a stroke. In addition, the utility after a non-fatal RTA, r_{nf} , is:

$$U(r_{nf}) = U_R$$

Then for each cycle (i) the expected utility of the population treated with T is:

$$U_i^{(T)} = \sum_X \left(q_i \left(X_j^{(T)} \right) U \left(X^{(T)} \right) \right) + [p(r_{nf}) \times q(r_{nf}) \times U(r_{nf})]$$

where $q(r_{nf})$ is the expected proportion of the cohort who could have a RTA, i.e. those in the cohort who are alive who have not had a disabling stroke:

$$q(r_{nf}) = 1 - q(\bar{d}) - s_{dis} [q(S) + q(pS)]$$

B.4.3 Costs

Similar to estimating utilities, each state is assigned a cost, $c \left(X^{(T)} \right)$. There is also a cost for fatal and a non-fatal RTA. The derivation of each of the costs is defined in Tables B.7 - B.10. Using the same notation as for calculating utilities the expected cost over the population for each cycle is calculated as:

$$C_i^{(T)} = \sum_X q_i(X^{(T)}) C(X^{(T)})$$

where:

$$\begin{aligned} C(O|a, T) &= C^{(T)} \\ \left. \begin{aligned} C(C|a, T) \\ C(pC|a, T) \end{aligned} \right\} &= C^{(T)} + C_C \\ \left. \begin{aligned} C(S|a, T) \\ C(pS|a, T) \end{aligned} \right\} &= C^{(T)} + C_S \\ C(RTA|a, T) &= C_R \\ C(D|a, T) &= 0 \end{aligned}$$

where $C^{(T)}$ is the cost of intervention T , C_C is the cost of a CHD events, C_S is the cost of a stroke, and C_R is the cost of a RTA.

B.5 Calculating the cost-effectiveness

The main outcomes of the CEA are the expected costs and QALYs over the lifetime of the cohort for each intervention. To obtain these the values from Sections B.4.2 and B.4.3 need to be aggregated over cycles for each intervention, T:

$$\bar{c}_T = \sum_{i=0}^{65} C_i^{(T)} (1 + disc_c)^{-i}$$

$$\bar{e}_T = \sum_{i=0}^{65} U_i^{(T)} (1 + disc_u)^{-i}$$

where $disc_c$ and $disc_u$ are the discount rates for costs and utilities respectively (Table B.2). The model follows individuals aged 50 at cycle $i = 0$ for a total of 65 years.

Using the principles set out in Section 1.1 this model is run K times each time taking a draw from the distributions of the uncertain parameters in Tables B.3 - B.11. Cost-effectiveness is calculated from the average costs and utilities over the K simulations (Section 1.1.1).

Appendix C

Search strategy for reviewing literature on adherence to MAD or CPAP as treatments for patients with OSAHS

This search strategy was presented in Sharples et al. (2014) [198]. This thesis has updated the search to all papers published before January 2015. The search strategy was:

Compliance search terms in Medline (January 2015)

1. exp Sleep Apnea Syndromes/
2. compliance.ti,ab.
3. adherence.ti,ab.
4. Continuous Positive Airway Pressure/
5. ("oral device" or "mad" or "mandibular advancement").mp
6. 2 or 3
7. 4 or 5
8. 1 and 6 and 7
9. limit 8 to (abstracts and english language and "review articles" and humans)
10. (long-term or long\$ term or (long adj3 term)).ti,ab.
11. 9 and 10

Appendix D

Reasons for inclusion and exclusion for papers reviewed for information on adherence to MAD or CPAP

Table D.1 Papers for inclusion or exclusion in the review of adherence to MAD or CPAP as treatments for OSAHS after review of full text - as at January 2015

Authors	Title	Reference	Reason sion/Exclusion	Inclu- for
Quinnell et al. [175]	Randomised controlled trial of mandibular advancement devices for obstructive sleep apnoea (TOMADO): One year follow-up	Conference Abstract - 22nd Congress of the European Sleep Research Society (2014)	No paper as yet but is the same population as in the case study so included	
Heeley et al.	Is long-term adherence to CPAP treatment different between stroke and cardiac patients with obstructive sleep apnoea (OSA)? Results of the SAVE trial	Conference abstract - 21st European Stroke Conference (2012)	Conference abstract	

Authors	Title	Reference	Reason for Inclusion/Exclusion	Inclu-
Singh PP.	A 7 year prospective study of mandibular anterior repositioning device in the management of obstructive sleep apnoea and snoring.	Conference Abstract - 27th Annual Meeting of the Associated Professional Sleep Societies (2013)	Conference abstract	
Vecchierini et al.	Efficacy and tolerance of a custom-made mandibular repositioning device for the treatment of obstructive sleep apnoea	Conference Abstract - 28th Annual Meeting of the Associated Professional Sleep Societies (2014)	Conference abstract	
Deloney et al.	The role of patient perceptions in compliance with continuous positive airway pressure	Conference Abstract - 25th Anniversary Meeting of the Associated Professional Sleep Societies (2011)	Conference abstract	
Furukawa et al. [77]	Long-term adherence to nasal continuous positive airway pressure therapy by hypertensive patients with pre-existing sleep apnoea	Journal of Cardiology 63 (4) (pp 281-285) April 2014	Included	
Lit et al.	Enhancing long-term continuous positive airway pressure (CPAP) compliance with a specialist nurse clinic	Conference Abstract: 18th Congress of the Asian Pacific Society of Respiration (2013)	Conference abstract	
Schoch et al.	Baseline predictors of adherence to positive airway pressure therapy for sleep apnoea: A 10 year single center observational cohort study	Respiration 87. (2) (pp121-128) 2014	Results presented graphically	

Authors	Title	Reference	Reason for Inclusion/Exclusion	Inclusion/Exclusion
Fordyce et al.	CPAP compliance - the first year and beyond	Conference Abstract - 5th International World Association of Sleep Medicine Congress and the 22nd Annual Congress of the Spanish Sleep Society (2013)	Conference abstract	
Van Zeller et al. [233]	5 years APAP adherence in OSA patients - do first impressions matter?	Respiratory Medicine 107 (12)(pp2046-2052) 2013	Included	
Chai-Coetzer et al.[35]	Predictors of long-term adherence to continuous positive airway pressure therapy in patients with obstructive sleep apnoea and cardiovascular disease in the SAVE study	Sleep. 36 (12) (pp1929-1937) 2013	Included	
Campos-Rodriguez et al.[32]	Long-term continuous positive pressure compliance in females with obstructive sleep apnoea	European Respiratory Journal 42 (5) (pp 1255-1262) (2013)	Included	
Araujo et al.	Retrospective study: 8 years of obstructive sleep apnoea analysis	Conference Abstract: 12th European Congress on Internal Medicine (2013)	Conference abstract	
Ag et al.	Long-term compliance, device longevity and effectiveness in use of "Mandibular Advancement Devices" for snoring and obstructive sleep apnoea	Conference Abstract: British Association of Oral and Maxillofacial Surgeons, BAOMS and Faculty of Dentistry 2013 Meeting	Conference abstract	
Salepci et al. [189]	CPAP adherence of patients with obstructive sleep apnoea	Respiratory Care 58(9) (pp 1467-1473) 2013	Time points of results were unclear	

Authors	Title	Reference	Reason for Inclusion/Exclusion	Inclu-
Tokunaga et al [228]	Long-term compliance with nasal continuous positive airway pressure therapy for sleep apnoea syndrome in an otorhinolaryngological office	European Archives of Oto-Rhino-Laryngology 270 (8) (pp 2267-2273) 2013	Included	
Brette et al.	A mandibular advancement device for the treatment of obstructive sleep apnoea: Long-term use and tolerance	International Orthodontics 10(5) (pp363-376) 2012	Unable to access paper	
Molony and Harper	Long-term compliance and device longevity in the use of mandibular advancement splints for snoring and sleep apnoea	Conference Abstract: 14th British Academic Conference in Otolaryngology, 2012	Conference abstract	
Chai-Coetzer et al.	Predictors of long-term adherence to continuous positive airway pressure therapy in patients with obstructive sleep apnoea and cardiovascular disease in the SAVE study	Conference Abstract: 24th ASM of Australasian Sleep Association and Australasian Sleep Technologists Association 2012	Conference Abstract for earlier Chai-Coetzer paper	
Sigurgunnarsdottir et al.	Continuous positive airway pressure and compliance - a long-term follow-up study: The Icelandic Sleep Apnoea Cohort (ISAC)	Conference Abstract: 21st Congress of the European Sleep Research Society 2012	Conference abstract	
Woehrle et al. [252]	Age and gender dependent adherence with continuous positive airway pressure therapy	Sleep Medicine 12 (10) (pp 1034-1035) 2011	No measure of continuation of treatment	
Heeley et al.	Is long-term adherence to CPAP treatment different between stroke and cardiac patients with obstructive sleep apnoea (OSA)? Results of the SAVE trial?	Conference Abstract: STROKE 2012	Replicate abstract	
La Piana et al.[121]	Long-term adherence to CPAP treatment in patients with obstructive sleep apnoea: Importance of educational program	Patient Preference and Adherence 5 (pp555-562) 2011	Included	

Authors	Title	Reference	Reason for Inclusion/Exclusion
Moor et al.	Reasons for the return of continues positive airway pressure machines in patients with obstructive sleep apnoea syndrome: Results from a UK district general hospital	Conference Abstract: American Thoracic Society Internal Conference ATS 2011	Conference abstract
Naksuk and Selcuk	Incidence of appropriate and inappropriate shocks in implantable cardioverter-defibrillator patients with obstructive sleep apnoea	Conference Abstract: 16th Annual Scientific Meeting of the Heart Failure Society of America	Conference abstract
Alves et al. [4]	Compliance to continuous positive airway pressure therapy in a group of Portuguese patients with obstructive sleep apnoea syndrome	Sleep and Breathing 16 (2) (pp555-562) 2012	Included
Stuck et al	Effects of continuous positive airway pressure on apnoea-hypophyseal index in obstructive sleep apnoea based on long-term compliance	Sleep and Breathing 1(2) (pp 467-471) 2012	No measure of treatment continuation
Schoch et al.	Long-term adherence to CPAP in sleep apnoea patients	Conference Abstract: 2011	Conference abstract
Vezina et al.[234]	Does propulsion mechanism influence the long-term side effects of oral appliances in the treatment of sleep disordered breathing?	Chest 140 (pp1184-1191) 2011	Included
Gagnadoux et al. [78]	Influence of marital status and employment status on long-term adherence with continuous positive airway pressure in sleep apnea patients	PLoS ONE 6(8) 211	Included
Galetke et al [79]	Long-term therapy with continuous positive airway pressure in obstructive sleep apnea: Adherence, side effects and predictors of withdrawal-a real life study	Respiration 82 (2) (pp155-161) 2011	Included
Barbe et al. [16]	Long-term effect of continuous positive airway pressure in hypertensive patients with sleep apnea	American Journal of Respiratory and Critical Care Medicine 181 (7) (pp 718-726) 2010	No measure reported of compliance in terms of how many discontinue treatment

Authors	Title	Reference	Reason for Inclusion/Exclusion	Inclusion/Exclusion
Aihara et al.	Long-term nasal continuous positive airway pressure treatment lowers blood pressure in patients with obstructive sleep apnea regardless of age	Hypertension Research. 33 (10) (pp 1025-1031) 2010	No measure reported of compliance in terms of how many discontinue treatment	
Kohler et al [119]	Predictors of long-term compliance with continuous positive airway pressure	Thorax 65 (9) (pp 829-832) 2010	Included	
Ghazal et al. [82]	A randomised prospective long-term study of two oral appliances for sleep apnoea treatment	Journal of Sleep Research 18 (3) (pp 321-328) 2009	Included	
Jauhar et al. [112]	Ten year follow-up of mandibular advancement devices for the management of snoring and sleep apnea	Journal of Prosthetic Dentistry 99(4) (pp 314-321) 2008	Included	
Gindre et al	Mandibular advancement for obstructive sleep apnea: Dose effect on apnea, long-term use and tolerance	Respiration 76 (4) (pp 386-392) 2008	Unable to access paper	
Campos-Rodriguez et al.	Long-term effect of continuous positive airway pressure on BP in patients with hypertension and sleep apnea	Chest 132 (6) (pp 1847-1852) 2007	No measure of continuous treatment effect	
Bromstrom et al. [28]	Association of type D personality to perceived side effects and adherence in CPAP treated patients with OSAS	Journal of Sleep Research 16 (4) (pp 439-447) 2007	Included	
Wietske et al	Acceptance and long-term compliance of nCPAP in obstructive sleep apnea	European Archives of Oto-Rhinolaryngology 264(9) (pp 1081-1086) 2007	Paper presents in terms of a range of follow-up times	
Marklund [132]	Predictors of long-term orthodontic side effects from mandibular advancement devices in patients with snoring and obstructive sleep apnea	American journal of orthodontics and dental orthopaedics 129 (2) (pp 214-221) 2006	Included	

Authors	Title	Reference	Reason for Inclusion/Exclusion	Inclusion/Exclusion
Chai et al [34]	Continuous positive airway pressure delivery interfaces for obstructive sleep apnoea	Cochrane Database of Systematic Reviews (4) 2006	Cochrane review - not included	
Sucena et al. [217]	Continuous positive airway pressure treatment for sleep apnoea: Compliance increases with time in continuing users	European Respirator Journal 27 (4) (pp 761-766) 2006	Only provides average number of minutes usage by year	
Chin et al [36]	Falls in blood pressure in patients with obstructive sleep apnoea after long-term nasal continuous positive airway pressure treatment	Journal of Hypertension 24 (10) (pp 2091-2099) 2006	Only provides average number of minutes usage	
Ayalon et al	Adherence to continuous positive airway pressure treatment in patients with alzheimer disease and obstructive sleep apnea	American Journal of Geriatric Psychiatry 14 (2) (pp 176-180) 2006	Patients have co-morbidities	
de Almeida et al.	Long-term compliance and side effects of oral appliances used for the treatment of snoring and obstructive sleep apnea syndrome	Journal of clinical sleep medicine 1(2) (pp 143-152) 2005	Unable to find paper	
Marin et al. [131]	Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: An observational study	Lancet 365 (9464) (pp1046-1053) 2005	No measure of continuation of treatment	
Johnson et al. [115]	Long-term continuous positive airway pressure (CPAP) outcomes from a sleep service using limited sleep studies and day case CPAP titration in the management of obstructive sleep apnoea/hypopnoea syndrome	Chronic respiratory disease 1(2) (pp83-84) 2004	Included	
Hollandt and Mahlerwein [102]	Nasal breathing and continuous positive airway pressure (CPAP) in patients with obstructive sleep apnea (OSA)	Sleep and Breathing 7 (2) (pp 87-93) 2003	No measure of timing in the paper	
McGown et al [136]	Long-term use of mandibular advancement splints for snoring and obstructive sleep apnoea: A questionnaire study	European Respiratory Journal 17 (3) (pp 462-466) 2001	Population with OSA and MAD smaller than 50	

Authors	Title	Reference	Reason for Inclusion/Exclusion	Inclu-
Loube	Oral appliance treatment for obstructive sleep apnea	Clinical Pulmonary Medicine 5(2) (pp124-128) 1998	Unable to find paper	
Kreiger et al	Long-term compliance with CPAP therapy in obstructive sleep apnea patients and in snorers	Sleep 19 (9 SUPPL) (pp S136-S143) 1996	Unable to find paper	
Pieters et al. [168]	Acceptance and long-term compliance with nCPAP in patients with obstructive sleep apnoea syndrome	European Respiratory Journal 9 (5) (pp 939-944) 1996	Only presents results by hours of usage	
Schmidt-Nowara et al	Oral appliances for the treatment of snoring and obstructive sleep apnea: A review	Sleep 18 (6) (pp501-510) 1995	Unable to find paper	
Krieger	Long-term compliance with nasal continuous positive airway pressure (CPAP) in obstructive sleep apnea patients and nonapneic snorers	Sleep 15 (6 SUPPL) (pp S42-S46) 1992	Results presented graphically (with only approximate numbers)	
Hoffstein et al.	Treatment of obstructive sleep apnea with nasal continuous positive airway pressure: Patient compliance, perception of benefits and side effects	American Review of Respiratory Disease 145 (4 I) (pp841-845) 1992	Unable to access find paper	
Waldhorn et al. [236]	Long-term compliance with nasal continuous positive airway pressure therapy of obstructive sleep apnea	Chest 97 (1) (pp 33-38) 1990	Included	